LR Foundation

Careful Industries

The AI Safety Paradox

A Literature Review on the Safe Adoption of Artificial Intelligence in Engineered Systems

Rachel Coldicutt OBE,
Uchenna Anyamele,
Madhuri Karak,
Dr. Odongo Oduor Joseph

February 2026

Lloyd's Register Foundation
Report Series No: 2026.1

About Lloyd’s Register Foundation

Our vision

Our vision is to be known worldwide as a leading supporter of engineering-related research, training and education, which makes a real difference in improving the safety of the critical infrastructure on which modern society relies. In support of this, we promote scientific excellence and act as a catalyst working with others to achieve maximum impact.

Lloyd’s Register Foundation charitable mission

- To secure for the benefit of the community high technical standards of design, manufacture, construction, maintenance, operation and performance for the purpose of enhancing the safety of life and property at sea, on land and in the air.
- The advancement of public education including within the transportation industries and any other engineering and technological disciplines.

About the Lloyd’s Register Foundation Report Series

The aim of this Report Series is to openly disseminate information about the work that is being supported by Lloyd’s Register Foundation. It is hoped that these reports will provide insights for research, policy and business communities and inform wider debate in society about the engineering safety-related challenges being investigated by the Foundation.

Lloyd’s Register Foundation is a Registered Charity (Reg. no. 1145988) and limited company (Reg. no. 7905861)
registered in England and Wales, and owner of Lloyd’s Register Group Limited.
Registered office: 71 Fenchurch Street, London EC3M 4BS, UK

T +44 20 7709 9166
E info@lrfoundation.org.uk
www.lrfoundation.org.uk

Contents

About this document

The purpose of this document is to inform the forthcoming Foresight Review into the Safe Adoption of AI in Engineered Systems by:

- exploring key themes related to AI adoption that have relevance for the future safe adoption of AI in engineered systems; and
- identifying useful areas for future investigation.

AI development and deployment is a fast-moving, complex field of practice that is influenced by technological breakthroughs, a competitive global marketplace, and conflicting international regulatory standards. Additionally, perspectives on the likely success of future developments in the field are still emergent and wide ranging.
This literature review surveys a significant body of relevant material across the fields of critical infrastructure, worker safety, and environmental safety to provide a clear starting point and evidence base for further foresight activities.

Methodology

This paper summarises findings from a literature review on the safe adoption of AI in engineered contexts. The majority of resources were searched during November 2025, with additional research taking place during December 2025 and January 2026. In recognition of the fast-moving and constantly changing evidence base, the literature search was time-boxed, taking 20 days in total. Literature was retrieved through a range of methodologies and the focus territories were India, the EU, China, the US, Kenya, and the UK.

The foundational themes for this literature review are infrastructure resilience, environmental safety, and worker safety. 333 sources were analysed, including established and peer-reviewed academic literature plus “grey” material such as news and media reporting and analysis, think tank reports, social media posts and newsletters, and pre-print research.

Credits

Writing and additional research:
Rachel Coldicutt

Worker Safety literature search and analysis:
Uchenna Anyamele

Environmental Safety literature search and analysis:
Madhuri Karak, PhD

Critical Infrastructure literature search and analysis:
Dr. Odongo Oduor Joseph

Careful Industries is a UK-based inclusive innovation studio. Through research, foresight, and prototyping we understand the impacts of technologies and create more inclusive futures through policy development and training.

www.careful.industries

Glossary

Algorithmic Bias: Systematic errors or prejudices embedded in AI systems that result in unfair or discriminatory outcomes, particularly affecting marginalized populations and vulnerable groups.

AI System: A machine-based system designed to operate with varying levels of autonomy that generates outputs such as predictions, recommendations, decisions, or content. An AI system typically comprises data, algorithms or models, and computational infrastructure working together to perform tasks with varying degrees of autonomy.

Artificial Intelligence (AI): Artificial intelligence is a broad, multi-disciplinary field of computer science that makes possible a number of advanced computing functions. These computer functions include analysing and processing data, using rules to organise and output information, and organising and completing tasks. Artificial intelligence systems also “learn” as they undertake these tasks, and so can make progress without requiring human intervention.

Assurance Frameworks: Structured methodologies and standards designed to validate and verify that AI systems meet safety, security, and performance requirements.

Autonomous Vehicles (AVs): Vehicles equipped with AI systems capable of perceiving their environment and making navigation decisions without human control.

Critical Infrastructure: Physical and digital systems essential to national security and public welfare, such as power grids, transportation networks, and water systems.

Generative AI: AI systems capable of creating new content, including text, images, and code, based on learned patterns from training data.

General Purpose AI (GPAI): An artificial intelligence system designed to perform a wide range of tasks across multiple domains, rather than being built for a single, narrowly defined function. The term has gained particular prominence in regulatory contexts, notably the EU AI Act, where it refers to AI models trained on broad data at scale (such as large language models) that can serve as a foundation for many downstream applications.

Hallucinations: Instances where AI language models generate false information, incorrect references, or fabricated content presented as factual.

Large Language Models (LLMs): Advanced AI systems trained on massive amounts of text data to understand and generate human language.

Model Collapse: The effect of training an AI model on the data it has generated, leading to decreasing performance over time.

Retrieval-Augmented Generation (RAG): The process of refining a large language model with additional information, taken from outside of its training sources.

Executive Summary

The safe adoption of general purpose AI in engineered systems is not a given. A highly competitive AI market that prioritises speed to market over rigorous testing, low levels of international regulatory alignment, and the tendency of large language model training methods to "guess" and "fake alignment" (Kalai et al. 2025, Greenblatt et al. 2024) are just some of the factors reshaping expectations around safety. As the promises and penalties of adopting frontier AI become apparent, the ability for industry leaders and policymakers to take informed decisions about technology adoption has never been more pressing.

This review takes a sociotechnical approach and our contention is that the safe adoption of AI **depends not just on the safe deployment of a given tool or system but on its safe creation and ongoing use.** As such, the material reviewed includes reports of emergent social impacts, technical analyses, and assessments of AI's impacts across the supply chain, focussing particularly on worker safety and environmental safety. The review surfaces multiple paradoxical outcomes created by the use of AI in complex environments and examines the disconnect between practical and applied advances in safety with the new risks and harms created by emerging and general purpose AI systems.

Overall, the literature shows that the AI safety paradox is already embedded as the "new normal" of modern technology development, anchored in place by the rapid adoption of advanced and general purpose AI, geopolitical turbulence and extraordinarily high levels of financial investment. In addition, the risk landscape is constantly evolving: through the period of this research, new risks and impacts related to the adoption of advanced and general purpose AI emerged on an almost day-to-day basis, shifting at a significantly different pace to existing standards and regulatory frameworks.

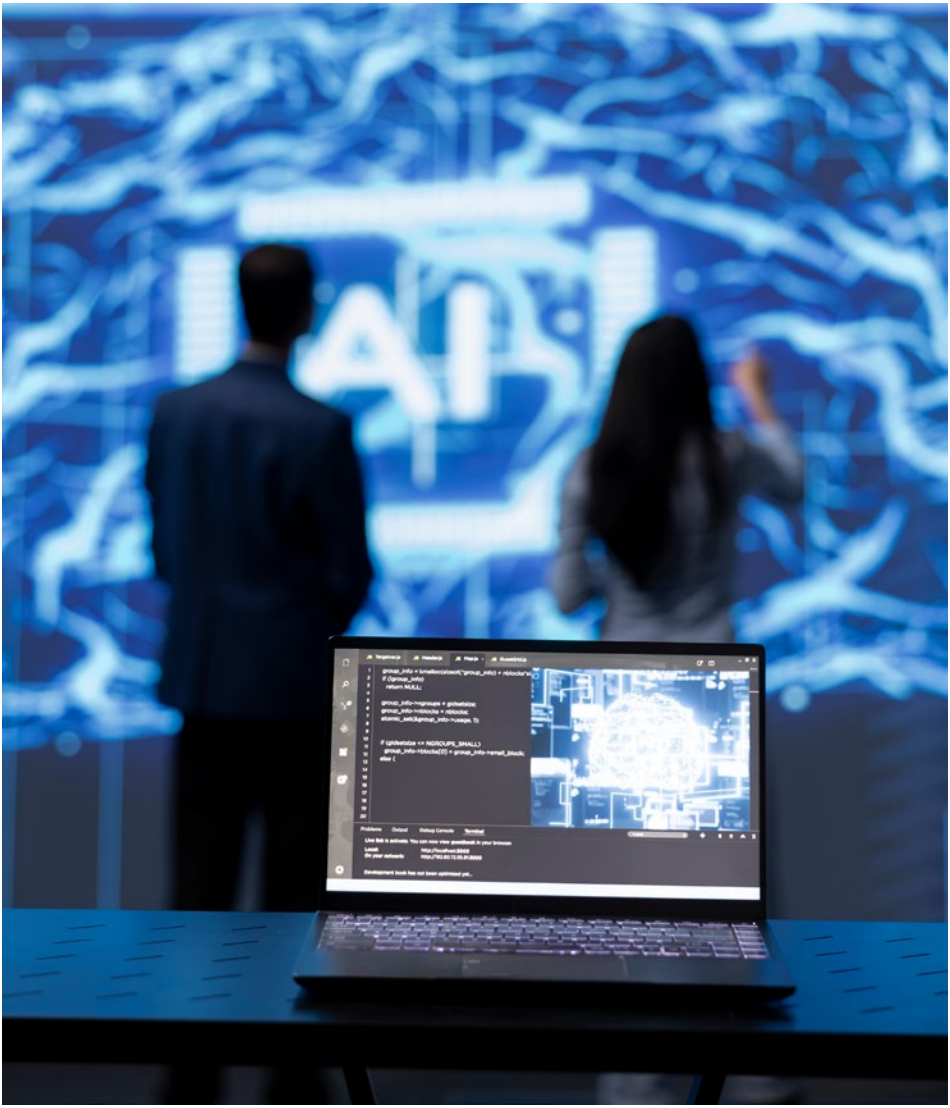
As such, practical interventions are required to ensure the further adoption of AI in engineered contexts is accompanied by meaningful risk assessments and useful due diligence; meanwhile, it also seems probable that alternative models of AI development are required to shift the balance in favour of a "safety-by-design" culture of innovation.

In particular this literature review finds that:

- The pressure to rapidly deploy emergent and untested technologies can displace both governance and assurance.
- The personal, social, and environmental impacts of AI development and deployment are disproportionately borne by those with the least power to shape its trajectory, including workers in Low and Middle-Income Countries, communities in climate-vulnerable regions, and populations subject to biased automated decision-making. These localised negative outcomes have an impact on social license for AI deployment, and may become more prevalent over time. Meaningful governance should take account of this power imbalance.
- The opacity of AI supply chains is a structural barrier to safety, and more transparency is required to enable the adoption of safer end-to-end systems.
- A sociotechnical view of the AI supply chain is required, including social and material components (human labour and natural resources) as well as technical aspects (data, model development, and compute). Additionally, the durational, post-deployment impacts of AI systems should be effectively monitored to understand and mitigate their overall impacts.

Overall, the findings of this review point to the fact that safe adoption of AI in engineered contexts requires better “end-to-end” models of assurance across the AI supply chain, supported by effective governance models that balance the competing interests of AI producers with those who are most affected by their outputs.

In the context of global regulatory misalignment and rapid technology adoption, these measures may seem ambitious; however, AI adoption across engineered systems is subject to significantly higher levels of scrutiny and assurance than either consumer or general-purpose enterprise technologies and is likely to afford the opportunity for alternative models of development and assurance to flourish.



```
group_info = kmalloc(sizeof(*group_info) + nblocks*si
if (!group_info)
  return NULL;

group_info->ngroups = gidsetsize;
group_info->nblocks = nblocks;
atomic_set(&group_info->usage, 1);


if (gidsetsize <= NGROUPS_SMALL)
  group_info->blocks[0] = group_info->small_block;
else {
```

```
Live link is activate. You can now view guestbook in your browser.
Local              http://localhost:3000
On your network    http://192.80.72.55.91:3000

Development book has not been optimized yet...
```

Introduction

This grey literature review examines the safety implications of AI deployment across critical infrastructure, worker safety, and environmental sustainability. The sources examined take a broad view across the AI life cycle, reflecting a sociotechnical approach to the AI supply chain, and considering the established and emerging uses and impacts of AI systems.

In this review, "general purpose AI" (GPAI) refers to AI systems that are not designed for a single task or domain but are capable of performing a wide range of distinct tasks across multiple applications. Large language models and foundation models are among the most prominent examples of general purpose AI, in contrast to narrow or task-specific AI systems built for defined purposes such as fault detection or fraud monitoring.

Shifting Contexts

The technologies surveyed in this review are at the nexus of several dimensions of change: at the time of writing, evolving technological capabilities, a volatile market, and geopolitical tensions are all shaping the near-future of AI adoption, leading to short-term outputs such as fluctuating hardware costs and rapidly changing regulatory factors. This review assumes that AI labs and leading firms will continue to develop and deploy general purpose AI systems over the coming years, and that general-purpose technologies will more frequently play some role in single-purpose systems. While the affordances of these systems may change over time, this review assumes that GPAI will continue to:

- be data-intensive;
- use natural materials to create hardware and natural resources to power storage, training and inference;
- be susceptible to hallucinations;
- and be adaptive in its outputs.

It is also worth noting that, in the field of AI, the concept of "safety" is disputed and evolving. As the UK AI Security Institute noted in 2023:

> *since our understanding of AI safety is nascent, it is not yet possible to build full safety cases that scale to risks posed by models significantly more advanced than those of today*

and there are ongoing disagreements as to whether the term “safety” relates to current societal harms or to future existential risks, with some critics venturing that the “safe” use and adoption of current forms of AI is not possible.

Rather than redefining the field of safety, this literature review examines the concept across three domains with a view to identifying emergent and future trends. Factors relating to these three areas — critical infrastructure, worker safety, and environmental safety — may arise individually or in combination at any given time.

The AI Safety Paradox

Across all areas of enquiry, the literature shows that the adoption of AI creates paradoxical outputs in which measurable safety improvements are often accompanied by new harms. These new harms may be disaggregated in appearance, perhaps taking the form of errors that are difficult to replicate or discern, or occur longitudinally as negative social impacts that roll out over time across a community. In some contexts, outcomes such as decreasing mental health outcomes for workers or increased miles travelled in autonomous vehicles may take much longer to materialise and be more difficult to quantify than any short-term safety gain caused by the use of predictive analytics or increased monitoring; in these cases traditional risk-management methods and approaches to cost–benefit analysis may not be appropriate to balance and compare outcomes. Moreover, the beneficiaries of AI adoption may not be the same people as those who are harmed: for instance, the data centres that power technologies that increase productivity in a workplace may also contribute to a shortage of potable water for a local community; the safety advantages of computer vision in a physically hazardous working environment may only be possible because low-paid and exploited workers in another location have labelled the underlying data sets. The balance of power that decides which of these outcomes is more important than others sits beyond traditional safety and assurance frameworks; however, addressing this disconnect and understanding the full range of impacts created by a technology or system is an essential aspect of safe deployment.

A Sociotechnical View of the AI Supply Chain

The opacity of the AI supply chain is a recurring theme across the literature.

A sociotechnical approach to AI safety requires transparency to be understood not only as a property of individual AI systems, but as a characteristic across the full AI supply chain. As such, a full end-to-end view of provenance and accountability is important; in addition to data and model explainability, human factors such as labour transparency and clarity on the use of natural resources and environmental costs should also be understood as components of the safe creation and operation of AI systems.

Taking a sociotechnical view recognises how each of these upstream processes shapes the safety, reliability, and fairness of downstream systems, and that technical explainability at the point of use cannot compensate for structural opacity at earlier stages of the lifecycle.

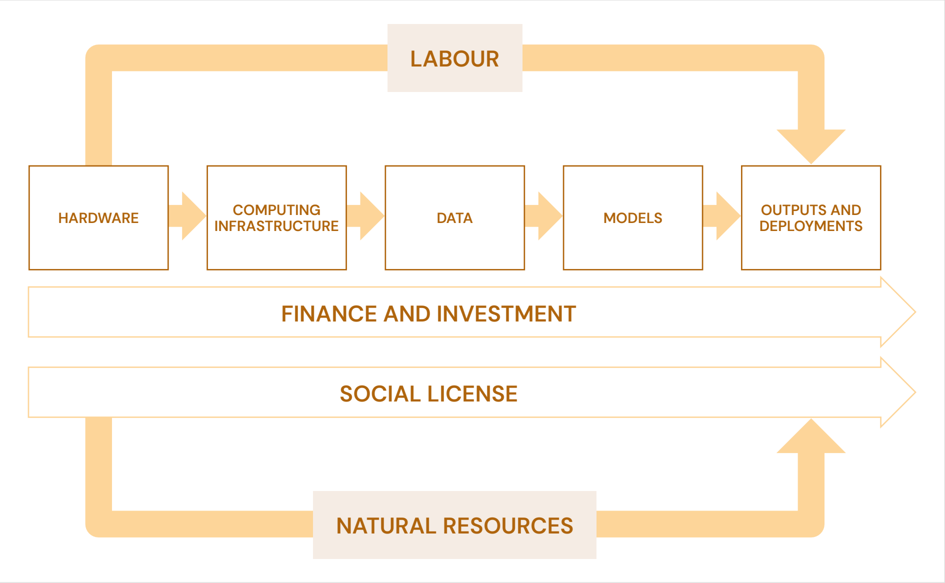


Figure 1 — A Sociotechnical View of the AI Supply Chain

Critical Infrastructure

AI systems are widely deployed across a range of critical infrastructure for a range of safety-related tasks such as fault detection, predictive maintenance, and resilience engineering; the introduction of advanced and general purpose AI into these systems creates a range of new safety risks that call for advanced forms of monitoring and assurance. However, the use of GPAI in critical infrastructure also brings into question whether reliable assurance is possible when reproducibility and transparency are low, hallucinations and "ersatz fluency" are established characteristics (Sood et al. 2025; Bender et al. 2021), and some models are capable of faking alignment in order to "pass" tests.

Worker Safety

The benefits of AI-driven automation are distributed unequally across the workforce. While human–robot collaboration can reduce physical risk in hazardous industries, it also introduces new workplace harms and can affect human dignity. Studies across China, the United States, and Germany document correlations between robot exposure and increased depression, substance abuse, and workforce disengagement. Meanwhile, the global infrastructure of data labelling and content moderation that underpins AI development relies on a largely invisible workforce operating under conditions that frequently violate international labour standards.

Environmental Safety

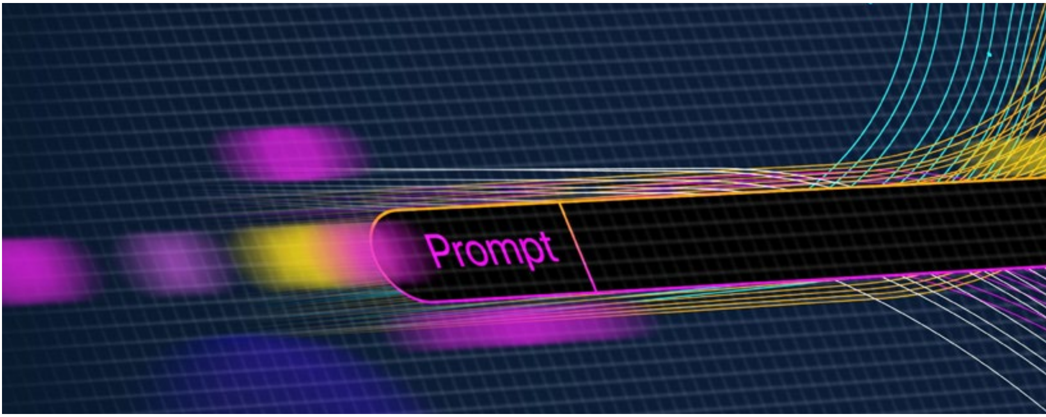
The environmental impacts of AI are substantial, unevenly distributed, and insufficiently measured. They span the full AI supply chain, from mineral extraction that destroys habitats to the disposal of electronic waste, and their burdens fall disproportionately on the people and habits of Low and Middle-Income Countries and on marginalised communities, while the lack of industry transparency on energy and water consumption means that neither policymakers nor affected communities have the information necessary to make informed decisions. Although alternative approaches — including sustainable AI frameworks, Frugal computing, and climate-positive applications — are under development, these remain marginal relative to the scale and pace of commercial expansion. Growing community opposition to data centres in the United States and Europe, alongside legal challenges in countries including Chile and Spain, indicates that the social licence for unchecked AI infrastructure expansion is narrowing.

Navigating Safety Trade-Offs

In aggregate, assessing the overall benefits of applied AI systems requires weighing up a range of trade offs that play out on a range of different scales. As the 2026 World Economic Forum (WEF) Global Risk Report points out:

Access to AI infrastructure as well as to electricity, internet access and data storage will amplify economic power shifts between countries over the next decade as AI’s productivity benefits bypass some populations entirely — albeit protecting them from some of the risks.

Good AI governance requires more than transparency and access to high-quality information; it also requires clear decision-making and high-quality accountability. However, establishing and maintaining equitable governance in a contested global field of development is perhaps more difficult than pursuing significant technological breakthroughs. As such, in parallel to better transparency and improved assurance and governance processes, the future safe adoption of AI in engineered systems also depends upon new technical approaches to unlocking frontier capabilities that do not reproduce the flaws of current approaches to general purpose AI.



1. Critical Infrastructure

A substantial body of literature addresses AI safety within critical-infrastructure systems including transport, financial services, critical manufacturing, and the energy, nuclear and water sectors. These systems are characterised by high interconnectedness and cascading failure risks, with risks and impacts having the potential to cause wide-ranging societal, economic, and environmental effects in addition to site-specific outcomes. This section explores themes in the literature relating to the efficacy of AI in selected areas of critical infrastructure.

AI systems are used in critical infrastructure systems for a range of purposes including fault detection, predictive maintenance, supply-chain optimisation, infrastructure safety, and resilience engineering; however, these systems can also introduce new failure modes (European Commission, Directorate-General for Migration and Home Affairs (DG HOME) 2025; Sarp et al. 2024; Tamascelli et al. 2024) and — as discussed throughout this literature review — also create a range of both upstream and downstream impacts that are often not visible to technically focussed assurance methodologies. Current approaches to assurance and risk assessment include both probabilistic and non-probabilistic models (Liu 2023; Gutfraind and Bier 2025; Kierans, Aidan; Rittichier, Kaley; Sonsayar, Utku; Ghosh, Avijit 2025; Wisakanto et al. 2025) and there is an established and growing body of applied practices and frameworks available to both developers and purchasing organisations (see, for instance, NIST 2021; York 2025). However, the literature shows that increasing use of advanced and general purpose AI technologies in critical-infrastructure systems is likely to increase the complexity and range of the risk landscape in ways that exceed the current capacities of assurance methods.

1.1 Some generic risks associated with advanced and general purpose AI technologies

a) The use of biased and faulty data sets

As mathematician David Spiegelhater says,

> *[W]hen we want to use data to draw broader conclusions about what is going on around us, then the quality of the data becomes paramount, and we need to be alert to the kind of systematic biases that can jeopardize the reliability of any claims. (Spiegelhater 2019)*

The use of large language models (LLMs), foundation models, and transformers that scrape content from the World Wide Web in the development of advanced and general purpose AI means that there is often little or no audit trail regarding the underlying contents or quality of data, particularly for more bespoke and tailored applications that are built "on top" of these systems. Birhane and McGann (2024) venture that the concept of "data completeness" in LLMs is fallacious, based on the assumption "that all of the essential characteristics can be represented in the datasets that are used to initialise and 'train' the model in question". Detailed audits of the underlying data set used in one market-leading LLM shows that sources are uneven and unrepresentative in scope (Kerche et al. 2026), and it is now well-established that available training data is disproportionately based on freely available English-language sources (Perez 2025; Knowledge 2022; Hao 2025). This means that critical systems can be dependent upon "faulty or biased" data sets, which undermine the foundational capabilities of the systems (Centre for Data Ethics and Innovation 2020).

b) AI-generated errors

Hallucinations, defined by IBM as "LLM… outputs that are nonsensical or altogether inaccurate", can create fabricated, inaccurate, or misleading information. Systems that are developed with or on top of advanced and general purpose AI systems are likely to have hallucinatory or confabulatory qualities and researchers at OpenAI have established that hallucinations are a feature of LLM development (Kalai et al. 2025). Errors due to hallucinations may be introduced at many different parts of the AI supply chain and their impacts may not only affect the operation of a system, but the quality of related information pertaining to a system's use and application, or wider public trust and understanding (World Economic Forum 2024a).

In the foundational paper "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", Bender et al. (2021) describe LLMs as having "ersatz fluency", creating text outputs that seem plausible but lack understanding. Workers may come into contact with hallucinations through a variety of routes, including simple search queries to verify or discover information or by using AI assistants to summarise material, create drafts documents, or generate code. AI-generated code may be integral to the underlying functionality and maintenance of advanced and general purpose AI systems or related applications, while AI-generated information may influence how a user operates or interacts with a system.

At the time of writing, the capabilities of commercially available tools for code generation have seen significant improvements, and the impacts of this for quality control and professional development are still playing out. Technologist Simon Willison emphasises that the most significant risk from LLM-generated code is likely to come from subtle bugs that are not visible to compilers, stressing that developers must continue to actively test and manually verify all code rather than trusting its appearance (Willison 2025).

Sood et al. argue that hallucinations — which manifest as false positives, false negatives, and erroneous threat interpretations — pose severe risks in critical systems and create new vulnerabilities to cyber attacks by causing alert fatigue, missed threats, wasted resources, and erosion of trust in automated security systems (Sood et al. 2025). While mitigation strategies for hallucinations do exist, including the integration of external knowledge systems like Retrieval-Augmented Generation (RAG) and human oversight, the "model collapse" hypothesis — which explains how LLMs models will collapse when trained recursively on AI-generated data (Shumailov et al. 2024) — points to another risk of over-reliance on LLMs; just as text hallucinations may not be immediately discernible, errors caused by model collapse may be plausible and undetectable in applications of generative AI or in the deployment of complex legacy systems.

Use of AI-enabled systems for information retrieval in the workplace is another emergent risk, with both the World Economic Forum and the UK AI Security Agency pointing to escalating levels of misinformation as a critical societal risk (AI Security Institute 2025; World Economic Forum 2024b). Research by the Tow Center for Journalism points to the fact that AI chatbots routinely return incorrect search results with a "high degree of confidence" which "presents users with a potentially dangerous illusion of reliability and accuracy" (Jaźwińska and Chandrasekar 2025), while an investigation by The Guardian has revealed that AI overviews generated in Google search results that appear to be authoritative can be "dangerous" and "alarming" (Gregory 2026). The World Economic Forum additionally flags the potential future reduction of workforce capability due to an over-reliance on AI as an emerging workplace risk (World Economic Forum 2026) while research by LLM-developer Anthropic points to the fact that over-reliance on AI assistants may lead to a decline in professional skills (Shen and Tamkin 2026).

c) Loss of control

In AI safety discourse, the concept of "loss of control" has previously been associated with future existential risks related to the development of Artificial General Intelligence (AGI) (Cent. AI Saf. 2023); the deployment of AI agents, which can independently perform non-sequential tasks within complex systems, makes more localised and discrete instances of loss of human control into a more immediately pressing issue. Bellogin et. al (2025) note that agentic AI systems:

> *may self-optimize toward unintended goals because of misaligned or poorly specified objectives. Failure modes include reward hacking, specification gaming, and goal mis-generalization (pursuing objectives that differ from the intended goal when faced with novel situations), which pose risks if left unchecked. Agentic AI models often remain internally opaque, complicating oversight even for their creators. This opacity complicates efforts to ensure alignment with human values and accountability in AI systems.*

In complex and safety-critical systems, the use of AI agents is considered by several experts to require multi-stakeholder, dynamic, and continuous governance, however the efficacy of such methods has not yet been robustly tested (Macapagal 2025; Government Digital Service 2025; World Economic Forum 2024b).

d) Cybersecurity risks

In the literature reviewed, increased vulnerability to cybersecurity attacks is a recurrent theme, with the UK AI Security Institute noting that "the same capabilities that could automate valuable work or reduce administrative burdens are inherently dual-use: they may also lower barriers for malicious actors" (AI Security Institute 2025).

Research by cybersecurity firm Darktrace shows that the "dual use" of AI, as both a technology that can both improve security monitoring and create new vulnerabilities, is front of mind for many security professionals, with 88% foreseeing that "AI cyberthreats will continue to have an impact on my organization in the future" (Darktrace 2025).

1.2 Case Studies

This section uses two case studies to demonstrate the complexity of the risk landscape created by existing use and deployment of AI deployment in applied domains.
The purpose of this is not to offer a full survey of uses of AI in safety-critical systems or across complex infrastructure, but to highlight useful "lessons learned".

1.2.1 Transport: Autonomous Vehicles

Autonomous vehicles (AVs) are found in a range of settings, including public transport, logistics, manufacturing, and private ownership. AVs rely on data that is labelled by human workers *(see Section 2, Worker Safety)*, often necessitate some degree of human-robot collaboration for safe operation, draw upon opaque decision-making systems, and increasingly make use of advanced and general purpose AI systems to deliver improvements in capabilities. Developments in AV assurance over the last decade demonstrate both the difficulties of accurate and replicable assurance methods and the importance of international regulatory alignment, while the potential that increased AV use will lead to overall higher emissions offers an example of the AI Safety Paradox.

Challenges of Assurance and Regulation

Trustworthy assurance for AVs is multi-modal and brings together cybersecurity, transparency, robustness, and fairness, as well as a range of domain-specific requirements including the regulation of automated lane-keeping systems and concerns relating to both pre-sale assurance and continuous monitoring (Fernández Llorca et al. 2025). Transparent assurance is further complicated by the use of opaque and proprietary Deep Neural Networks (Utesch et al. 2020).

AVs also create challenges for criminal negligence and liability: identifying who is responsible and to what extent when an AV is involved in an accident or other illegal incident, establishing whether users or operators have sufficient information to foresee these failures, and the potential limits of genuine human control or oversight all rely upon system transparency, which is not always a given (Giannini and Kwik 2023).

Internationally, there is also a landscape of regulatory and legislative approaches. While, for instance, territories including the EU, Japan, China, and Singapore have agreed uniform type approval requirements for automated driving functions, the United States continues to operate a self-certification model for manufacturers (UNECE 2017; NHTSA, n.d.; UNECE 2021). Broadly, there is agreement on the need for multi-modal safety validation that combines simulation, controlled testing, real-world validation, and ongoing operational monitoring, but enforcement of this regimen is uneven.

This lack of international alignment creates the conditions for a safety gap in which different standards are applied to AVs than to other high-risk AI systems, and also fails to establish a precedent for other kinds of human–robot collaboration. Dunphy (2024) posits that this novel complexity makes the regulation of AVs a "wicked problem" that requires collaborative and adaptive governance.

Public opinion

AVs have a more tangible public profile than many other kinds of automation in critical systems. To date, autonomous taxis have had a key role in shaping public opinion on AVs, with a study by Park et al. (2025) showing that an overall high-level of public concern about high-risk use cases was tempered by the utility of increased convenience and affordability *(see below, The Rebound Effect)*. Qualitative research by Liu et al. (2025) shows that visible safety features are important factors in passenger trust, while earlier public research in China (Liu et al. 2018) highlights the importance of "perceived improvements in traffic efficiency — such as reduced congestion, shorter travel times, and enhanced travel experiences" in creating public approval.

Meanwhile, in the US, high-profile accidents have served to impact trust levels with specific manufacturers (Contreras 2025). For example, the safety record of AV manufacturer Tesla has led to the US National Highways and Transport Safety Administration (NHTSA) opening an investigation covering 2.88 million Tesla vehicles after receiving 58 reports of traffic safety violations (O'Kane 2025). This was followed by a ruling by a California administrative law December 2025 that Tesla's use of "Full Self-Driving" (FSD) is "unambiguously false and counterfactual" and that "Autopilot" represents an "unlawful tradition of intentionally using ambiguity to mislead consumers while maintaining some level of deniability" (Dow 2025). A survey of 8000 American consumers found that "48% of consumers believe Tesla's FSD technology should be illegal, and [that] FSD puts off more potential Tesla purchasers than it attracts by more than two-to-one" (EVIR 2025).

The Rebound Effect

The “rebound effect” of autonomous vehicles refers to the phenomenon where efficiency gains from AV technology paradoxically lead to increased vehicle miles traveled (VMT), due to the increased ease and decreased cost of travel. Research reveals a consistent positive rebound effect where improved fuel efficiency (0–40%) and lower travel time costs (up to 45%) increase VMT, partially offsetting expected energy savings, with rural households exhibiting more substantial rebound effects due to longer travel distances (Letmathe and Paegert 2025; Ahn et al. 2025; Massar et al. 2021).

When examining the entire life cycle, autonomous electric vehicles might emit 8% more greenhouse gas emissions on average compared to non-autonomous electric vehicles due to these effects, demonstrating that without proactive policies, widespread AV adoption is likely to induce a rise in VMT (Onat et al. 2023).

Conclusions

The safe adoption of AVs relies not only on accurate assurance but on the development of holistic policy measures that relate to their use and deployment, as well continuous safety monitoring and assessment. AVs differ from some other uses of AI discussed here in that they are not only used in industrial and infrastructural settings but are also used by private individuals, public bodies, and businesses for day-to-day transport and service delivery; this general-purpose use sets a higher bar for complexity and safety than is needed in some other cyberphysical systems, but it also provides a useful basis for considering complex future scenarios for other kinds of automation. Even in territories with established regulatory guardrails, the development of AVs does not proceed on an established on a “safety-by-design” basis, meaning that accurate assurance and safety testing will not be a given in any further new developments.

1.2.2 Case Study: Financial Services

The use of AI across the financial services sector creates significantly improved fraud detection and predictive analytics capabilities, while also giving rise to new privacy and security concerns and increased potential for biased automated decisions that deepen social divisions and create new kinds of social and economic exclusion (Eubanks 2017; O'Neill 2017). Externally, the use of AI tools by bad actors creates new security vulnerabilities for both individuals and institutions (Bank of England 2025), while the impacts of AI-powered systems and tools across the financial services sector offers an example of both the kind of trade-offs that take place in new technology adoption and the new threat surfaces that arise through adoption elsewhere.

Uses of AI systems

AI systems have a number of uses across the financial services sector including risk management, high-speed stock-market trading, credit evaluation, fraud protection, and customer-service delivery. These tools are often developed outside of the financial sector, with the Bank of England flagging the risk of dependency on a small number of providers (2025).

Examples of improved AI-driven capabilities include improved detection of major network failures across financial macro-systems (Ren et al. 2021); effective prediction of unexpected financial business events (Liu 2025); and the Adjusted Gross Granular Model, developed as a tool to predict microfinance institution failures in Low and Middle-Income Countries, enabling regulators to take early action against financial instability in susceptible communities (Garcia-Lopez et al. 2025).

Additionally, Sharma and Priya (2025) demonstrate how AI-based FinTech systems can contribute to improved access to financial-management skills and services for digitally included people in North India, but — showing the paradoxical nature of AI systems — use of these tools also creates a competing risk of introducing algorithmic bias in risk-management and decision-making systems. In the UK, a review by the Financial Conduct Authority identified several sources of bias in supervised machine learning across the financial services sector, leading to unfair or discriminatory outcomes for people from protected or vulnerable groups (Bogiatzis-Gibbons et al. 2024).

New vulnerabilities and new safeguards

Financial services are vulnerable to the hostile use of AI tools by bad actors to create both systemic, large-scale vulnerabilities and targeted instances of fraud and deception. Saha et al (2025) highlight the extent to which AI is being used to infiltrate financial systems and perpetrate consumer fraud, with generative AI being used for phishing, deepfake fraud, and adversarial attacks; in the UK, Cifas note that reported cases of financial fraud rose by 50,000 in 2025, with the easy availability of AI services and "fraud toolkits" noted as a key enabler (Cifas 2025). In the first half of 2025, the United States, United Kingdom, and Canada were the countries most exposed to data breaches, with high-impact cybersecurity attacks creating privacy risks for regulators, financial institutions, and millions of customers (Experian 2026).

New governance needs

Overall, the use of AI systems creates new governance pressures: in Australia research indicates that licensees are adopting AI technologies faster than they are updating their risk and compliance frameworks (Australian Securities and Investments Commission 2024). Across Asia-Pacific, 98% of financial institutions have had to increase their compliance and anti-fraud operations, at an overall cost $45 billion, in response to estimated losses of $688 billion (Allianz 2026; Thapar 2025).

Conclusions

The use of AI systems both inside the financial services sector and by bad actors outside of the sector shows the wide range of potential impacts within complex data-driven systems. Within organisations, AI creates the safety paradox observed throughout in this literature review, improving safety capabilities while simultaneously creating new risks; externally, use of AI systems create new critical vulnerabilities, which are often met with internal deployments of AI that improve monitoring and protection. Overall this is a circular model of impact that demonstrates the range of interconnected threat surfaces that complex, data-rich systems are subject to. Separately, the potential societal harms of biased and faulty decision making in financial systems mirror deployment issues of AI systems in domains such as justice and welfare ('Royal Commission into the Robodebt Scheme' 2023; Guo, 2025.; Adams Bhatti 2025), pointing to failures in system design and assurance practices across the supply chain.

1.4 Conclusions

While AI technologies deliver measurable improvements in efficiency, safety monitoring, and predictive capability across critical infrastructure systems, they also introduce new failure modes, vulnerabilities, and governance challenges that are often poorly understood and unevenly regulated.

The generic risks outlined above — biased and incomplete training data, hallucinations and AI-generated errors, loss of human control in agentic systems, and expanded cybersecurity attack surfaces — are well-established in the literature but are not always visible at the point of purchase or deployment, and can be compounded by the complexity and interconnectedness of infrastructural systems. For instance, in transport, the challenges of assurance, liability, and international regulatory alignment coexist with impacts on public trust and rebound effects that undermine the environmental case for adoption; in financial services, AI-driven improvements in fraud detection and predictive analytics are counterbalanced by AI-enabled threats from external actors, creating a circular dynamic in which the technology is both the problem and the proposed solution.

2. Worker Safety

A significant cluster of research addresses the human dimensions of AI adoption, showing that AI safety extends beyond technical systems to organisational and worker wellbeing. Worker safety is another paradoxical issue: while in some instances the use of AI systems can reduce physical harm, the production of these systems can also entail exploitative working conditions and their deployment can lead in some cases to increased physical risk or rising worker stress and low-workplace satisfaction, giving rise to numerous societal, psychological, and economic risks.

The literature emphasises that safe AI adoption requires attention to worker rights and safety protocols, worker transition support and reskilling opportunities, and equitable labour market policy (Abeliansky and Beulmann 2019; Peng et al. 2025; Adriana 2024; Nelson et al. 2023; Khurram et al. 2025). Additionally, as detailed in Section 1: Critical Infrastructure, there is also emergent potential for AI-generated errors to create new kinds of workplace harms.

2.1 Human–Robot Collaboration

Globally, robotics are used to tackle productivity, safety, and labour-shortage issues in a range of industries including construction, manufacturing, and hospitality, and are often used for high-risk, physically intensive tasks. Use of robotics produces considerable physical safety benefits to human workers by preventing traumatic injuries, minimising the adverse health effects of hazardous conditions and repetitive physical actions, and reducing safety issues that arise from human error. However, emerging findings show that human–robot collaboration can also introduce new workplace and mental-health risks for workers.

China is the world’s largest user of workplace robotics. Collaborative robots (also called “cobots”) support human staff by handling repetitive, physically demanding tasks and their use has improved the physical safety of workers in heavy industrial contexts and created new employment opportunities for “cobot operators”. Meanwhile, small-to-medium enterprises responsible for over 60% of GDP and 80% of urban employment in China are making increasing use of cobots to compensate for human labour shortages (Gihleb et al. 2022; The State Council Information Office, People’s Republic of China 2025).

Robot exposure is, however, also correlated with increased mental stress. Studies on the influence of robot adoption on the mental health of the Chinese working-age population show that the deployment of robotics can increase depression and anxiety for workers, with notable impacts on male workers who are concerned about their current and future income and employment opportunities (Zou and Chen 2025, Liu et al. 2024). These findings are supported by studies on workers’ mental health in the US which show that robot penetration leads to considerable increases in drug or alcohol-related deaths and other mental health problems, including anxiety. In Germany, however, there is some evidence that an overall increase in labour-market and retraining opportunities has lessened the mental-health impacts of robotic deployments for some workers, although these concerns remained high among those who considered their skillsets to be “at risk” (Abeliansky and Beulmann 2019; Gihleb et al. 2022).

In complex working environments such as construction there is evidence that specific new physical risks can arise because of the inability of robotic systems to handle dynamic workplaces; this unpredictability necessitates additional worker training, site safety protocols, and ergonomic design (Peng et al., 2025). Unlike factory-oriented industries, most construction robots operate in close proximity to workers. Moreover, design considerations in the programming of construction robots can be extremely complicated due to the presence of various large or significant entities (such as equipment, material, or people) in frequently reconfigured working conditions, and while mitigations for this are possible, their effective deployment requires a high-degree of accurate human monitoring and surveillance to be effective.

Additionally, the physical and psychological uncertainty that arises from human collaboration with robots in complex settings also gives rise to increased levels of human stress, with evidence that "following" a robot in the workplace is a high-intensity human task that in turn reduces the levels of interest and satisfaction to be gained from employment (Liu 2023; Simone et al. 2022). A "robot-first" approach to human–robot collaboration can also lead to an overall degradation in working conditions, including increased noise and unpleasant lighting conditions. In workplaces that rely on low-skilled labour, there is evidence that a focus on efficiency can also lead to the normalisation of severely reduced quality working conditions. For instance, in corporate retail warehouses where robot "pickers" take the lead, there are numerous reports from India, the US, and the UK of human working conditions being deprioritised in favour of productivity measures, leading to a removal of basic dignities such as bathroom breaks and the increased likelihood of repetitive and other physical injuries (OSHA, 2023.; Bloodworth 2019; Banerji 2021). Additionally, increased workplace monitoring may increase stress for workers, with UK trade union representatives speaking to the "dehumanising effect" of invasive tracking and related unfair disciplinary measures that some logistics workers now experience, currently without recourse to legal challenge (Labour Res. Dep. 2025).

While these impacts may be discrete, automation-induced mental health issues also have societal and economic impacts, with low workplace morale leading to higher staff turnover rates, increased absenteeism, and lower productivity (Ali, 2015). In addition to decreased levels of personal satisfaction, which may be a factor in broader societal unrest and decreasing societal trust, Gallup's 2023 report State of the Global Workplace shows disengaged workers cost $8.8 trillion in lost productivity (Gallup, 2023).

2.2 Ghost Work: Data Labelling, Content Moderation, and "Automated" Decisions

AI requires significant quantities of labelled, categorised, moderated, and annotated data. Due to the low transparency of many AI models, the provenance of underlying data sets is often uncertain or unknown. As such, automated processes built on top of LLMs or Foundation Models may draw upon data from an unknown range of sources, meaning that provenance can be difficult, or sometimes impossible, to ascertain (Zewe 2024). This lack of transparency means that organisations procuring or implementing AI-enabled tools and services are likely to have little visibility over the workers who have shaped the underlying data sets.

Investigations by academics, journalists, and civil society organisations have shown that human labour is often central to the creation of data-driven services and high-quality structured data sets, but it is not always visible to the end-user or purchaser, reduced to the status of "Ghost Work" (Gray and Suri, 2019). Computer vision, for instance, is frequently augmented or made possible by human labelling. In the case of data for autonomous vehicles, this labour includes detailed and repetitive tasks such as creating 2D and 3D bounding boxes around objects including vehicles and pedestrians; labelling every pixel in an image to identify drivable surfaces, lanes, and road boundaries; tracing lane markings and marking specific object features to help the vehicle navigate and predict movement; and labelling 3D maps from LiDAR sensors for precise distance measurements (Macgence, n.d.). Investigations into these industries have uncovered social and environmental concerns regarding working conditions of the AI workforce in countries including the USA, Mexico, the Philippines, India, and Kenya. Meanwhile some innovations that appear to be fully automated are almost entirely people powered: for instance Amazon's "Just Walk Out" technology, which enables frictionless shopping experiences, is delivered by a team of 1000 human workers in India who manually review the majority of transactions (Bitter, 2024).

Human data labelling has been commonplace since Amazon Mechanical Turk (also known as mTurk) launched in 2005 as a crowdsourcing marketplace to distribute "human intelligence tasks" to a globally dispersed network of remote workers. Named after an 18th-century chess automaton that concealed a human operator, the platform exemplifies how AI development relies on human labour: workers earn a median wage of approximately $2 per hour to complete tasks such as transcribing audio and labelling images. While mTurk has become crucial infrastructure for machine learning — workers build and label the training datasets that enable AI systems to learn — the platform has also become a case study in labour exploitation, with workers depersonalised through numeric anonymity and lack of meaningful communication with task requesters (Schwartz 2019).

In 2019, Gray and Suri estimated that at least 8% of American workers had participated in the data-driven "ghost economy", often working with few benefits or protections and usually earning less than legal minimums for traditional work (Gray and Suri 2019). Sarah T. Roberts' ethnographic study of global social-media content moderators sets out how the escalating volumes of global social-media content have given rise to US companies outsourcing low-paid, high-stress data cleaning and moderation tasks to workers in Low and Middle Income countries including Mexico and the Philippines, where workers occupy "colonies of exploitation", often with no or few psychological or employment protections and no mechanisms for redress (Roberts 2019).

More recent research by human rights and labour organisation Equidem, based on 113 interviews with data labellers and content moderators across several countries, exposes the extreme occupational, psychological, sexual, and economic harms faced by those moderating violent content and training AI models for major Silicon Valley platforms. The report documents mental-health abuses and violations of international labour standards, including International Labour Organisation (ILO) protections on fair wages, the right to unionise, and safeguards against forced labour (Scroll. Click. Suffer, 2025)

A number of Chinese technology companies operate data labelling practices in Kenya, hiring workers through informal networks of WhatsApp groups and paying as little as $5.42 per day for up to 12 hours of work. These working practices prioritise speed and reduced costs over wellbeing, and workers have no formal contracts, minimal labour protections, and no knowledge of which companies employ them, a model that critics describe as "digital colonialism". Kenya's high youth unemployment rate (67% as of July 2025) and weak labour protections make it an attractive hub for international employers; however, the Kenyan government is currently drafting regulations to protect vulnerable workers (Dosunmu and Waithira 2025).

2.3 Conclusions

Across the full spectrum of AI deployment, the benefits of automation are uneven, with increased safety for some workers and consumers being underwritten by physical, psychological, and economic harms to others.

In instances of human–robot collaboration, the potential for reduced physical risk coexists with documented increases in worker stress, anxiety, and degraded working conditions, particularly where efficiency-driven “robot-first” approaches deprioritise human dignity and wellbeing. These impacts are not merely individual but carry measurable societal and economic costs, from increased substance abuse and mental health crises to financial losses caused by workforce disengagement.

Meanwhile, the global infrastructure of data labelling and content moderation that underpins AI development relies on a largely invisible workforce operating under conditions that frequently violate international labour standards, with workers in Low and Middle-Income Countries disproportionately bearing the psychological and economic costs of training systems from which they derive little benefit. The opacity of AI supply chains compounds these harms by obscuring the human labour embedded within ostensibly automated systems, making it difficult for procuring organisations to exercise due diligence and for regulators to enforce protections. These findings point to the need for governance and assurance frameworks that integrate worker rights, supply chain transparency, and labour market protections as core components of responsible AI adoption. Without such measures, the expansion of AI systems will entrench a model of development in which safety gains for end users are systematically subsidised by the exploitation and diminished wellbeing of the workers who build, train, and operate alongside these technologies.

3. Environmental Safety

The rapid expansion of AI technologies presents yet another paradox — this time for environmental sustainability and ecological protection. While AI can offer powerful tools for addressing climate change and environmental monitoring, its deployment simultaneously introduces significant environmental risks through resource extraction and ecosystem disruption.

The environmental impacts of AI are distributed across the AI supply chain and include rare-earth minerals mining, chip manufacturing, the building of data centres, and energy and water use related to data storage, training and inference, and the disposal of e-waste (UNEP 2024). These impacts are generated through significantly different processes, are spatially disaggregated, and take place across different national regulatory regimes, so there is no universally accepted assessment framework and no standard for end-to-end impact measurement.

The intersection of AI and climate change has received increasing attention from researchers and policymakers in recent years. The UN Environmental Protection Agency and International Science Council's "Global foresight report on potential environmental impacts of AI" (2024) provides systematic assessment of environmental risks that are likely to accelerate significantly without policy intervention. Meanwhile, there is a growing movement of global organisations working to address AI's environmental impacts through practical actions and policy influencing (Green Screen Coalition et al. 2025). These remedies include investigations into imposing limits on the extent of AI development and the development of alternative methods, including small language models, on-device learning, open-weight models and Frugal AI (High-Level Summary of the AI Act | EU Artificial Intelligence Act, n.d.; Varoquaux et al. 2025; Willison, n.d.; Gandikota 2025; Mozilla.ai 2025).

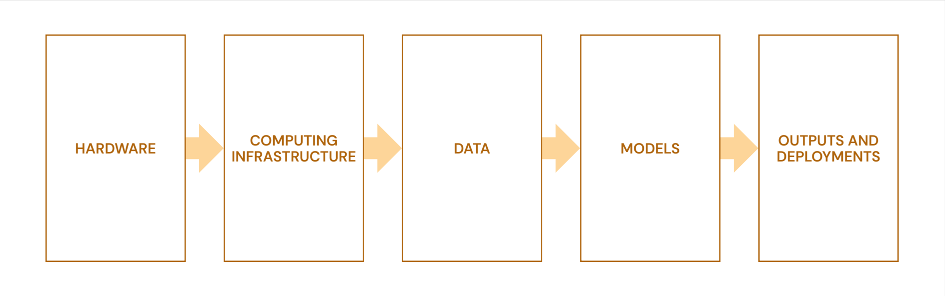


Figure 2 — Components of the AI supply chain with significant environmental impact

Although there is growing public awareness of the local environmental impacts of data centres in the US *(see Section 3.2)*, the environmental impacts of AI development and deployment are disproportionately experienced in Low and Middle-Income Countries countries *(see Sections 3.1, 3.3)*. Critiques of this analyse how existing governance frameworks can often reinforce global inequalities (UNEP 2024, Regattieri 2025), with more recent scholarship emphasising the importance of participatory governance and sociotechnical measurement frameworks *(see Sections 3.4, 3.5)*. At the heart of environmental governance for AI is a tension between economic assessments that favour industrial outcomes for High Income Countries and more holistic assessments that prioritise the needs of indigenous communities and of Low and MIddle-Income Countries.

3.1 Material Extraction

AI systems depend on rare earth elements, lithium, cobalt, and other critical minerals essential for hardware manufacturing. Securing critical materials for AI infrastructure is becoming a central geopolitical concern, and there are both economic and environmental implications as nations compete for access to limited mineral resources (Howey, 2023), with national industrial policy frameworks that promote AI frequently taking priority over environmental protections (AI Now Institute, 2024).

The extraction of minerals is associated with environmental damage including habitat destruction, water pollution, and biodiversity loss (Garofalo et al., 2025), and current approaches to securing materials for AI may accelerate global environmental destruction as High-Income Countries extract materials from LMICs (FERN, 2023). Lithium mining, which is marketed as essential for green technologies, creates particular harm in water-scarce regions: for instance, in Chile's Puna de Atacama, it has led to downstream exposure to chemicals for local communities, with significant negative outcomes for people's health and livelihoods (Blair et al., 2023). This highlights the inherent tension between pursuing "green" AI technologies and the environmental costs of acquiring the materials necessary to build them. Meanwhile McQue et al. (2025) address how "global critical minerals" affect AI development pathways and environmental outcomes. This research documents the resource constraints that will increasingly limit AI expansion if environmental externalities continue to be ignored.

3.2 Energy Consumption and Carbon Footprint

Geographic variation in environmental impacts reflects differences in electricity grid composition, cooling requirements related to climate zones, and varying levels of environmental regulation across jurisdictions. Some data centre locations in cooler climates benefit from lower cooling energy demands, while those in warmer regions face elevated environmental costs. This geographic variation suggests that environmental impacts of AI deployment are not uniformly distributed, with certain regions bearing disproportionate environmental burdens from hosting global AI infrastructure.

Large-scale AI systems, particularly LLMs and the data centres that train and run these models, consume significant amounts of electricity. Reporting by NPR (Kerr 2024) highlights significant increases in the carbon footprints of both Google and Microsoft related to the use of AI, with Google reporting that:

In 2023, our total GHG emissions were 14.3 million tCO2e, representing a 13% year-overyear increase and a 48% increase compared to our 2019 target base year. This result was primarily due to increases in data center energy consumption and supply chain emissions. As we further integrate AI into our products, reducing emissions may be challenging due to increasing energy demands from the greater intensity of AI compute, and the emissions associated with the expected increases in our technical infrastructure investment. (Google 2024)

This energy demand directly translates into greenhouse gas emissions, particularly in regions where electricity generation relies on fossil fuels. Overall, the AI boom means that "the climate footprint of the cloud is growing when it should be shrinking" as more physical infrastructure is built out (Walton et al., 2024). This can affect local residents in a number of ways, including increased electricity bills and intense pressure on the electricity grid, leading in some cases to power cuts and instability of the local electricity supply (Gooding 2025, O'Neill 2025).

Advanced and general purpose AI technologies such as transformer-based LLMs, deep learning systems, and generative AI applications typically require orders of magnitude more electricity than traditional software, particularly when deployed for high-throughput applications. Some estimates suggest that AI-driven solutions consume 10–15 times more energy than conventional approaches to similar tasks, raising questions about the environmental cost-benefit analysis of AI adoption in energy-intensive sectors (Bashir et al. 2024).

Much of the literature points to the fact that there is insufficient transparency from AI companies regarding their energy use, a problem exacerbated by the fact that the majority of AI research now takes place within technology companies, where findings are subject to corporate publication policies (Eastwood 2023). This lack of accurate data has led to public disagreements between leading figures in AI over the accuracy and availability of environmental and water-use data (Taft 2025). While there is not consistent global monitoring, some patterns are beginning to emerge: globally, 43% of data centres operate in areas of high water stress, with India and Australia particularly exposed (Standard and Poor Global 2025). Our literature search also highlighted a growing body of news coverage relating to the localised impacts on potable water of data centre construction and operation. For example, in the US, in 2022 a public water emergency was declared in Morrow County, Oregon after recirculated water used by hyperscale data centres was shown to be polluted, while in Mansfield, Georgia the operation of a 50-acre data centre increased water use by 200 million gallons per year, with nearby residents suffering severe contamination of water wells and low water pressure (Cooper 2025; Steingraber et al. 2025). In Aragon in Spain, water used by three planned data centres is forecast to use 755,720 cubic metres of water annually, enough to irrigate more than 570 acres of farmland. With 75% of Spain at risk of desertification, the combination of climate change and data centre expansion has led the campaign group Tu Nube Seca Mi Río ("Your cloud is drying my river") to call for a moratorium on new data centres (SourceMaterial 2025; Abdullahi 2025).

O'Donnell and Crownhart (2025) have conducted empirical analysis of AI's energy footprint, providing detailed calculations of the electricity required for training and deploying AI models at scale. Their research shows that, while individual AI queries use modest amounts of electricity — ranging from 114 joules for small text models to over 3 million joules for high-quality video generation — the massive scale of deployment, combined with insufficient industry transparency, poses significant climate concerns, and restrains accurate forward planning. O'Donnell and Crownhart's research indicates that 80%–90% of AI's energy use takes place during inference, when models make predictions, rather than in training, and so energy use is likely to increase as adoption continues. (On this, see also UNEP 2024, which notes that "Numerous institutions and research groups are working to develop AI frameworks that minimize the use of training data and optimize model operation during the inference stage to reduce compute and storage needs and address sustainability goals.").

In some regions across the United States, the impacts of data centres have become politically charged, with local residents campaigning against their construction and ongoing operation. The establishment of the xAI supercomputer Colossus in Memphis, Tennessee in 2024 has been a case in point: operating 33 methane-powered gas turbines, local campaign groups estimate that the supercomputer's operation has increased smog in the city by 30–60%. As Brabenec (2025) points out, Colossus is, "located in a poor, predominantly Black Memphis community with historically high rates of pollution-related illness and disproportionate rates of industrial pollutants." (See also Southern Environmental Law Center 2025). Writing for environmental magazine Heatmap in early 2026, Holzman estimates that at least 25 data centre projects were cancelled in the US due to community action in the previous year, with almost 100 more proposed building projects facing opposition.

O'Neill (2025) outlines the challenge for Ireland, where the data centre sector contributes €7.3 billion to the economy and consumes 21% of the country's electricity, a figure projected to rise to 32% by 2026. This surge in demand threatens to create supply shortfalls and grid strain, particularly in Dublin, and to exceed Ireland's legally binding carbon budgets, potentially by 77–114 million tonnes of CO_2 equivalent in the period 2026–2030. The Irish Commission for Regulation Utilities has established new connection policies requiring data centres to source 80% of power from renewable energy sources and provide on-site storage or generation capacity, while broader infrastructure investments aim to support renewable energy deployment, electrification, and circular economy initiatives like waste heat recovery for district heating networks. This challenge of balancing economic, environmental, and regulatory demands is being experienced in other European countries, and could ultimately lead to a significant slow down in roll-out and development.

3.3 Sustainable AI Development and Environmental Justice

The concept of “sustainable AI” has emerged as researchers attempt to develop AI systems that minimise environmental impact while delivering utility. Van Wynsberghe (2021) proposes frameworks for ensuring that AI development and deployment align with sustainability principles and argues for fundamental rethinking of how AI systems are designed, trained, and deployed to reduce environmental footprints while maintaining beneficial applications. Raman et al. (2024) note that a wide-variety of practices make up publications in the wider field of “green and sustainable” AI, identifying three significant clusters: advances in Green AI for energy optimisation; responsible AI for sustainable development; and big-data driven computational advances. This lack of clarity on the meaning of the term “sustainable AI” may present a barrier to its future achievement.

The ongoing debate over whether AI can be a tool for attaining the Sustainable Development Goals hinges on whether new and effective methods for developing “sustainable AI” can become a reality (Hilliger et al. 2025). Research on digital technologies and climate resilience indicates that, while AI may enhance climate adaptation in some applications, the net environmental benefit depends heavily on the energy sources powering AI infrastructure, and the specific use cases being supported. In regions relying on fossil fuel-based electricity generation, the environmental costs of AI deployment may offset or exceed the climate benefits of the systems themselves (Argyroudis et al., 2022). Multistakeholder work by the Green Screen Coalition (2025) offers concrete policy recommendations for constraining AI development within planetary boundaries, recommending limits are imposed on AI deployment.

Reporting from the 2025 United Nations Climate Change Conference (known as Cop30) in November 2025, Oliver Milman explored the differing views of attendees, capturing in broad terms the two sides of the AI and environmental safety argument. While some interviewees saw AI use as a frippery that is largely “guzzling energy for slop content”, others expressed hope that AI technologies might lead to breakthroughs in essential capabilities such as more accurate weather forecasting, pointing to the launch of the UN AI Climate Institute, “a new global initiative aimed at fostering AI ‘as a tool of empowerment’ in developing countries to help them tackle environmental problems”. However, the literature indicates that poor international governance creates a barrier to AI becoming a reliable tool for delivering climate justice, with Lehuedé (2022) highlighting how AI infrastructure development frequently occurs without proper consideration of impacts on indigenous territories and lands managed by local communities.

Climate positive uses of AI include environmental monitoring and climate science across weather and climate forecasting; disaster prevention and early warning; tracking and reducing pollution; carbon neutrality and clean energy; fashion industry sustainability; and implementation in agricultural and food systems. Many of these systems are still in development, although some are actively in use on the ground: for example, the MyAnga app sends precision weather forecasts to Kenyan pastoralists, enabling herders to plan ahead and can save time looking for green pastures (UN 2025).

3.4 Environmental Impact Assessment and Measurement

The literature indicates that environmental impact assessments need to expand beyond carbon emissions and energy and water use to include other societal and environmental impacts across the AI lifecycle.

Kneese (2024) identifies methodological limitations in how environmental impacts of AI are currently quantified and proposes the inclusion of empirical studies that measure and develop standards, while Dominguez Hernández et al. (2024) propose frameworks for understanding how AI affects multiple levels of environmental systems simultaneously, including individual behavioral impacts, social-level consequences, and biospheric effects.

Additionally, Nonnecke and Dawson (2022) address “Human rights impact assessments for AI”, proposing that human rights frameworks can be extended to address environmental impacts, with the concept of “environmental rights” serving as a foundation for protecting ecosystems from AI-related damage. Hosseini et al. (2025) propose that generative AI should be evaluated through integrated social-environmental frameworks that assess impacts on vulnerable communities and ecosystems simultaneously.

3.5 Conclusions

The environmental impacts of AI development and deployment represent the clearest illustration of the AI safety paradox: a technology promoted as a tool for addressing climate change and environmental degradation is simultaneously accelerating both.

Across the AI supply chain environmental costs are substantial, unevenly distributed, and insufficiently measured. The evidence shows that these burdens fall disproportionately on Low and Middle-Income Countries and on marginalised communities, whether through mineral extraction that destroys habitats and contaminates water sources, or through the siting of energy-intensive infrastructure in areas already subject to environmental stress and pollution-related illness. The lack of industry transparency on energy and water consumption, combined with the absence of standardised end-to-end impact measurement, means that policymakers nor affected communities currently have the information necessary to make informed decisions about AI infrastructure.

While work is underway on alternative models including sustainable AI frameworks, Frugal computing approaches, and climate-positive applications, the literature suggests that these efforts remain marginal relative to the scale and pace of expansion driven by commercial imperatives. Crucially, the concept of "sustainable AI" itself lacks definitional clarity, and the net environmental benefit of AI deployment depends on variables, including energy source, geographic context, and specific use case, that are rarely assessed holistically.

Growing community opposition to data centres in the US and Europe, alongside legal challenges in countries including Chile and Spain, indicates that the social licence for unchecked AI infrastructure expansion is narrowing. Taken together, this points to the need for good governance, with a growing movement for imposing enforceable limits on AI's environmental footprint. Additionally, the literature points to a need for full lifecycle impact assessments that centre the rights of affected communities, particularly indigenous peoples and those in climate-vulnerable regions, and address the structural inequalities created by AI deployments.

4. AI Governance and Regulatory Frameworks

This section does not give a thorough overview of international legislative frameworks, but explores whether the issues raised elsewhere in the literature review — including the need for a sociotechnical approach to the AI safety paradox and the lack of transparency across the AI supply chain — are resolved through current approaches.

4.1 The Need for Multi-Modal Regulatory Alignment

Over the last decade, the tension between the pace of legislative progress and the speed of technological development has been frequently mooted as the most significant sticking point for AI regulation (Greenstein and Zamboni 2025; Wheeler 2023; Zaidan and Ibrahim 2024). However, rapidly splintering international approaches and geopolitical tensions have more recently come to the fore, suggesting that a globally fragmented regulatory landscape (perhaps along the lines of the “Four Internets” described by O’Hara and Hall in 2021) may be the most decisive point of friction in coming years. These regulatory frictions have also seen the rise of self-regulatory mechanisms, including the growth of industry-specific AI assurance as a discipline.

The pace of development in advanced and general purpose AI in the first half of the 2020s tends to also suggest that competition — geopolitically, between the United States and China and between individual firms (Hao 2025; Vieira 2025) — has a profound effect on the cadence at which products are released to the market. Additionally, the perspective of individual investors and investment firms shapes the availability of capital and infrastructural investment (Kariuki 2025). As such, multiple aspects of alignment — political, economic, and technical — are needed between influential territories to meaningfully shape the safe release of advanced and general purpose AI-driven products and services to market.

4.2 EU Focus

The European Union's Artificial Intelligence Act takes a safety-first approach, requiring pre-deployment testing and human oversight for high-risk applications and providing a voluntary Code of Practice for general purpose AI which, at the time of writing, has been co-signed by 28 AI development companies (Article 60; Article 26; European Commission, 2025). Recital 55 also specifically addresses the use of AI in fields including critical digital infrastructure management, road traffic control, and water, gas, heating and electricity supply systems. Additionally, the Critical Entities Resilience (CER) Directive is a standardised system for cybersecurity and critical entity protection across eleven sectors. (DIRECTIVE (EU) 2022/2555 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on Measures for a High Common Level of Cybersecurity across the Union, Amending Regulation (EU) 2022).

The 2025 Paris AI Action Summit appeared to mark a strategic pivot for the EU, with the €200 billion InvestAI initiative showing a new focus on pursuing European AI sovereignty and industrial competitiveness, including investment in "national champions" to reduce dependence on US technology (Giambertoni 2025; Scott 2025; Hernández 2025).

4.3 US Focus

The United States' approach to AI regulation is characterised by an increasingly assertive executive federal deregulatory agenda, a raft of voluntary frameworks, and a rapid proliferation of state-level laws, leading to unevenness across the national regulatory landscape (White & Case 2025; Baker Botts 2026; Congressional Research Service 2025). The National Institute for Science and Technology (NIST) AI Risk Management Framework, published in January 2023, provides a voluntary, non-sector-specific framework for organisations to manage AI-related risks (NIST 2023) while the Department of Homeland Security Framework offers voluntary guidance for AI supply chain responsibility (US Department of Homeland Security 2024).

At the time of writing, the approach is shaped by a sequence of executive orders, with most recent — "Ensuring a National Framework for Artificial Intelligence" (December 2025) — issued with the intention of creating a "minimally burdensome" unified regulatory approach to sustain and enhance US global AI dominance (White House 2025; America's AI Action Plan 2025; White House 2025a). Meanwhile, analysis of US Security and Exchange Commission (SEC) Form 10-K filings demonstrates growing corporate acknowledgment of AI-related risks, though disclosure practices remain inconsistent (Marin et al. 2025).

4.4 China Focus

China's approach to AI regulation combines ambitious national industrial strategy with increasingly granular, technology-specific regulation. The "New Generation Artificial Intelligence Development Plan" (NAIDP), released in 2017, sets out China's ambition to become a global leader in AI by 2030, driven by substantial resource investment and minimal regulation (Carnegie Endowment for International Peace 2025). The country aims to achieve 70% AI penetration in key sectors by 2027 and 90% by 2030, with a vision of building a fully AI-powered economy and society by 2035 (Li 2025).

China has also supplemented its regulations with an extensive standards infrastructure, in line with the overall approach for specific regulatory frameworks. These include:

- The first version of the National Technical Committee 260 on Cybersecurity's AI Safety Governance Framework (September 2024), which acknowledges that AI "presents significant opportunities to the world while posing various risks and challenges." The Framework outlines principles for AI safety governance, classifies anticipated risks, identifies technological measures to mitigate those risks, and provides governance measures and safety guidelines (Tobey et al. 2024).
- A National Information Security Standardisation Technical Committee consultation on the "Artificial Intelligence Safety Standard System (V1.0)" (February 2025), with the intention of establishing comprehensive safety guidelines for AI development and application. Key areas of concern include model security, data privacy, bias mitigation and the ethical deployment of AI systems (Global Legal Insights 2025)
- In April 2025, the State Administration for Market Regulation and the Standardisation Administration jointly released three national standards: a Generative AI Data Annotation Security Specification, a Security Specification for Pre-training and Fine-tuning Data, and Basic Security Requirements for Generative AI Services (White and Case LLP 2025).

4.5 Conclusions

The three jurisdictional approaches outlined above reveal fundamentally divergent regulatory attitudes and competing visions for AI's role in economic and societal development. The EU pivot towards sovereignty-driven investment alongside its safety-first regulatory framework operates on a different footing to the globally competitive approaches of China and the US, while the underlying differences in technical approaches speaks to a lack of alignment in granular, technical approaches.

This divergence poses significant challenges for the alignment needed to enable safe advanced and general purpose AI deployment, particularly as competitive pressures between nations and firms continue to accelerate product release cycles. Moreover, none of these frameworks analysed for this literature review address the supply chain transparency concerns addressed in sections 2 and 3, or fully embrace the need for a sociotechnical understanding of AI impacts. While multistakeholder initiatives are in place, the fundamentally differing nature of these underlying approaches speaks to the fact that top-down, global regulatory alignment is unlikely to provide any short-term solutions for the safe adoption of AI.

5. Conclusion

The relationship between AI and safety is fundamentally paradoxical. In each domain examined, AI systems simultaneously deliver measurable improvements and introduce new, often poorly understood risks.

The AI safety paradox operates at multiple scales. In critical infrastructure, AI-driven improvements in fault detection, fraud prevention, and predictive maintenance coexist with new failure modes created by biased training data, hallucinations, loss of human control in agentic systems, and expanded cybersecurity attack surfaces. In the workplace, automation that reduces physical risk for some workers is underwritten by the exploitation and psychological harm of others — from warehouse operatives whose working conditions are degraded by robot-first efficiency models to data labellers and content moderators in Low and Middle-Income Countries who bear the hidden human costs of training AI systems. Environmentally, a technology promoted as a tool for addressing climate change is accelerating resource extraction, energy consumption, and ecological damage, with the burdens falling disproportionately on marginalised communities and climate-vulnerable regions. In each case, the technology is positioned as both the source of the problem and the proposed solution, creating circular dynamics that resist straightforward resolution.

The literature does not present a straightforward solution to the AI safety paradox, but points to a number of specific challenges:

1) The opacity of AI supply chains is a structural barrier to safety.

The lack of transparency over training data provenance, the invisibility of human labour in ostensibly automated systems, and the absence of standardised environmental impact measurement mean that organisations procuring or deploying AI-enabled tools frequently lack the information necessary to conduct meaningful risk assessment or due diligence. This opacity is not incidental; it is a feature of commercial AI development, where proprietary systems, corporate publication policies, and fragmented global supply chains combine to obscure the full range of upstream and downstream impacts.

2) Existing assurance and governance frameworks are insufficient to address the sociotechnical complexity of AI safety.

The literature consistently shows that technically focussed assurance methodologies fail to capture the broader societal, environmental, and labour impacts that accompany AI deployment. Risk frameworks that treat AI safety as a primarily technical problem miss the systemic effects documented across the domains reviewed here: the rebound effects that undermine the environmental case for autonomous vehicles, the mental health impacts of human–robot collaboration, or the ways algorithmic bias in financial services deepen existing social divisions. A sociotechnical approach that integrates technical, social, environmental, and economic dimensions would improve safety assurance by offering a more complete picture of the end-to-end impacts of adopting a given system.

3) The international regulatory landscape is fragmented in ways that compound rather than mitigate risk.

The fundamentally divergent approaches of the European Union, United States, and China reflect competing visions of AI's role in economic and societal development. None of the frameworks examined fully address supply chain transparency, worker protections, or environmental justice, and the competitive pressures between nations and firms continue to accelerate product release cycles in ways that outpace regulatory capacity. Top-down global alignment appears unlikely in the short term, suggesting that alternative mechanisms — including sectoral standards, procurement requirements, and multi-stakeholder governance initiatives — will need to play a more prominent role.

4) The distribution of AI's costs and benefits is profoundly unequal.

Across every domain reviewed, the harms of AI development and deployment are disproportionately borne by those with the least power to shape its trajectory: workers in Low and Middle-Income Countries who label data and moderate content under exploitative conditions; communities in climate-vulnerable regions whose water sources are contaminated by mineral extraction or whose air quality is degraded by data centre operations; and populations subject to biased automated decision-making in domains such as finance, welfare, and justice. The literature makes clear that AI safety cannot be meaningfully assessed without attending to these distributional questions, and that governance frameworks which fail to centre the rights of affected communities will entrench rather than address existing inequalities.

5) The pressure to rapidly deploy emergent and untested technologies can displace both governance and assurance.

The generic risks of general purpose AI identified in Section One of this review are well-documented in the literature but mitigations for them are not adequately addressed in testing frameworks or in the solutions available to regulators, procurers, or affected communities. The growing body of evidence on environmental, labour, and societal harms has not so far served to limit the scope of GPAI development; as such, without the development of new technical models or a pivot to a "safety-by-design" approach, the continued expansion of GPAI systems will likely continue to generate safety deficits that are displaced onto the most vulnerable.

Full Bibliography

“A Short History of Jobs and Automation.” *World Economic Forum*, 3 Sept. 2020, https://www.weforum.org/stories/2020/09/short-history-jobs-automation/.

Abdullahi, Aminu. “AI Data Centers Boom Is Draining Water From Drought-Prone Areas.” *TechRepublic*, 9 May 2025, https://www.techrepublic.com/article/news-ai-data-centers-drought/.

Abeliansky, Ana Lucia, and Matthias Beulmann. *Are They Coming for Us? Industrial Robots and the Mental Health of Workers*. 2019.

Adams Bhatti, Sophia. *AI in Our Justice System*. Jan. 2025.

Adriana, Parvu. *Digitalization of Work and Its Impact on Worker Safety and Health*. 2024. *ResearchGate*, https://doi.org/10.1007/978-3-031-54671-6_2.

Agarwal, Vibhor, et al. “CodeMirage: Hallucinations in Code Generated by Large Language Models.” arXiv:2408.08333, arXiv, 8 July 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2408.08333.

Ahn, Kyoungho, et al. *The Rebound Effect of Autonomous Vehicles on Vehicle Miles Traveled: A Synthesis of Drivers, Impacts, and Policy Implications*. Nov. 2025. *vtechworks.lib.vt.edu*, https://hdl.handle.net/10919/139779.

AI Advisory Body. *Governing AI for Humanity: Final Report*. United Nations, 2024, https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

AI and Data Labelling: “I Felt like My Life Ended.” www.bbc.com, https://www.bbc.com/news/av/world-africa-66514287. Accessed 10 Nov. 2025.

AI IS NOT OK | Union Action Needed on Artificial Intelligence (AI) at Work. 20 Jan. 2025, https://www.ituc-csi.org/ai-is-not-ok-union-action-needed.

AI Now Institute. *AI Nationalism(s): Global Industrial Policy Approaches to AI*. 12 Mar. 2024, https://ainowinstitute.org/publications/research/ai-nationalisms-global-industrial-policy-approaches-to-ai.

“AI Risks That Could Lead to Catastrophe | CAIS.” *Center for AI Safety*, https://safe.ai/ai-risk. Accessed 24 Jan. 2026.

AI Security Institute. “Frontier AI Trends Report PDF - The AI Security Institute (AISI).” *AI Security Institute*, Dec. 2025, https://www.aisi.gov.uk/frontier-ai-trends-report/pdf.

“AI-Definitions-HAI.Pdf.” https://hai-production.s3.amazonaws.com/files/2020-09/AI-Definitions-HAI.pdf. Accessed 1 Dec. 2025.

“AI-in-Our-Justice-System-Final-Report.Pdf.” https://files.justice.org.uk/wp-content/uploads/2025/01/29201845/AI-in-our-Justice-System-final-report.pdf. Accessed 25 Jan. 2026.

Alexander Obaigbena, et al. “AI and Human-Robot Interaction: A Review of Recent Advances and Challenges.” *GSC Advanced Research and Reviews*, vol. 18, no. 2, Feb. 2024, pp. 321–30. *DOI.org (Crossref)*, https://doi.org/10.30574/gscarr.2024.18.2.0070.

Alexander, Robert, et al. “Engineering Safety-Critical Complex Systems.” *CoSMoS 2008: Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation*, Sept. 2028.

Ali, Vikki. *The High Cost of Low Morale—And What To Do About It*. 2015, https://www.trinet.com/insights/high-cost-low-morale-what-to-do-about-it.

Allianz. *Allianz Risk Barometer: Identifying the Major Business Risks for 2026*. Jan. 2026.

Alsaigh, Roba, et al. “AI Explainability and Governance in Smart Energy Systems: A Review.” *Frontiers in Energy Research*, vol. 11, Jan. 2023. *Frontiers*, https://doi.org/10.3389/fenrg.2023.1071291.

Amironesei, Razvan. *Assessing Risks and Impacts of AI (ARIA): Pilot Evaluation Report*. NIST AI NIST AI 700-2, National Institute of Standards and Technology, 2025, p. NIST AI NIST AI 700-2. *DOI.org (Crossref)*, https://doi.org/10.6028/NIST.AI.700-2.

Ananny, Mike, and Kate Crawford. “Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability.” *New Media & Society*, vol. 20, no. 3, Mar. 2018, pp. 973–89. *SAGE Journals*, https://doi.org/10.1177/1461444816676645.

Andresz, Sylvain, et al. “Artificial Intelligence and Radiation Protection. A Game Changer or an Update?” *arXiv.Org*, 9 June 2023, https://doi.org/10.1051/radiopro/2022004.

Araszkiewicz, Michał; Nalepa, Grzegorz J.; Pałosz, Radosław. "The Artificial Intelligence Act. Taking Normative Imbalances Seriously | Internet Policy Review." Online Open-Access Journal (News/Opinion). *Internet Policy Review*, 2024, https://policyreview.info/articles/news/artificial-intelligence-act-taking-normative-imbalances-seriously/1817.

Argyroudis, Sotirios A., et al. "Digital Technologies Can Enhance Climate Resilience of Critical Infrastructure." *Climate Risk Management*, vol. 35, Jan. 2022, p. 100387. *ScienceDirect*, https://doi.org/10.1016/j.crm.2021.100387.

Article 3: Definitions | EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/3/. Accessed 1 Dec. 2025.

Artificial Intelligence, Platform Work and Gender Equality | European Institute for Gender Equality. 9 Dec. 2021, https://eige.europa.eu/publications-resources/publications/artificial-intelligence-platform-work-and-gender-equality?language_content_entity=en.

Ashrafi, Negin, et al. "AI-Driven Solutions to Improve Safety and Health: Application of the REDECA Framework for Agricultural Tractor Drivers." *PLOS Global Public Health*, vol. 5, no. 6, June 2025, p. e0003543. *PLoS Journals*, https://doi.org/10.1371/journal.pgph.0003543.

"Australia: Kenyan Data Labellers Make Modern Slavery Allegations against AI Company Appen." *Business & Human Rights Resource Centre*, https://www.business-humanrights.org/en/latest-news/australia-kenyan-data-labellers-make-modern-slavery-allegations-against-ai-company-appen/. Accessed 10 Nov. 2025.

Australian Securities and Investments Commission. *Beware the Gap: Governance Arrangements in the Face of AI Innovation*. ASIC, 2024.

Baker Botts. "U.S. Artificial Intelligence Law Update: Navigating the Evolving State and Federal Regulatory Landscape | Thought Leadership." *Baker Botts*, Jan. 2026, https://www.bakerbotts.com/thought-leadership/publications/2026/january/us-ai-law-update.

Banerji, Oishika. "Unsafe Working Conditions at Amazon Warehouse and Its Impact on Employees : An Insight." *iPleaders*, 27 July 2021, https://blog.ipleaders.in/unsafe-working-conditions-amazon-warehouse-impact-employees-insight/.

Bank of England. *Financial Stability in Focus: Artificial Intelligence in the Financial System*. 2 Dec. 2025, https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025.

Barez, Fazl, et al. “Open Problems in Machine Unlearning for AI Safety.” *arXiv.Org*, 9 Jan. 2025, https://arxiv.org/abs/2501.04952v1.

Bashir, Noman, et al. “The Climate and Sustainability Implications of Generative AI.” *An MIT Exploration of Generative AI*, Mar. 2024. *mit-genai.pubpub.org*, https://mit-genai.pubpub.org/pub/8ulgrckc/release/2.

Becker, Adam. *More Everything Forever: AI Overlords, Space Empires, and Silicon Valley’s Crusade to Control the Fate of Humanity*. Basic Books, 2025.

Bellogín, Alejandro, et al. *Systemic Risks Associated with Agentic AI: A Policy Brief*. Oct. 2025.

Belu, Andreea. “Twin Transition: From Green and Digital towards Defense and Security.” *Green Web Foundation*, 19 Aug. 2024, https://www.thegreenwebfoundation.org/news/twin-transition-from-green-and-digital-towards-defense-and-security/.

Bender, Emily M., et al. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? □.” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* [Virtual Event Canada], 2021, pp. 610–23. *DOI.org (Crossref)*, https://doi.org/10.1145/3442188.3445922.

Bender, Emily M., and Alex Hanna. *The AI Con: How to Fight Big Tech’s Hype and Create the Future We Want*. Bodley Head, 2025.

“Better Accountability Mechanisms Can Help Protect Society’s Most Vulnerable from AI-Based Harms.” *The Alan Turing Institute*, https://www.turing.ac.uk/blog/better-accountability-mechanisms-can-help-protect-societys-most-vulnerable-ai-based-harms. Accessed 27 Jan. 2026.

Bipartisan Policy Center. “What Past Waves of Automation Can Teach Us About AI.” 9 July 2024, https://bipartisanpolicy.org/article/what-past-waves-of-automation-can-teach-us-about-ai/.

Birhane, Abeba, and Marek McGann. “Large Models of What? Mistaking Engineering Achievements for the Human Linguistic Agency.” *Language Sciences*, vol. 106, Nov. 2024, p. 101672. *ScienceDirect*, https://doi.org/10.1016/j.langsci.2024.101672.

Bitter, Alex. “Amazon’s Just Walk Out Technology Relies on Hundreds of Workers in India Watching You Shop.” *Business Insider*, https://www.businessinsider.com/amazons-just-walk-out-actually-1-000-people-in-india-2024-4. Accessed 18 Jan. 2026.

Blair, James J. A., et al. “The ‘Alterlives’ of Green Extractivism: Lithium Mining and Exhausted Ecologies in the Atacama Desert.” *Revue Internationale de Politique de Développement*, no. 16, Apr. 2023. *DOI.org (Crossref)*, https://doi.org/10.4000/poldev.5284.

Bloodworth, James. *Hired: Six Months Undercover in Low-Wage Britain*. Atlantic Books, 2019.

Bogiatzis-Gibbons, Daniel, et al. *Research Note: A Literature Review on Bias in Supervised Machine Learning*. Dec. 2024.

Booth, Robert, and Robert Booth UK technology editor. “More than 140 Kenya Facebook Moderators Diagnosed with Severe PTSD.” *The Guardian*, 18 Dec. 2024. Media. *The Guardian*, https://www.theguardian.com/media/2024/dec/18/kenya-facebook-moderators-sue-after-diagnoses-of-severe-ptsd.

Borboni, Alberto, et al. “The Expanding Role of Artificial Intelligence in Collaborative Robots for Industrial Applications: A Systematic Review of Recent Works.” *Machines*, vol. 11, no. 1, Jan. 2023, p. 111. *www.mdpi.com*, https://doi.org/10.3390/machines11010111.

Bouchikhi, Maeva El, et al. “The Internet of Things Deployed for Occupational Health and Safety Purposes: A Qualitative Study of Opportunities and Ethical Issues.” *PLOS ONE*, vol. 19, no. 12, Dec. 2024, p. e0315671. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0315671.

Braband, Jens, and Hendrik Schäbe. “On Safety Assessment of Artificial Intelligence.” arXiv:2003.00260, arXiv, 29 Feb. 2020. *arXiv.org*, https://doi.org/10.48550/arXiv.2003.00260.

Brabenec, Ren. “A Billionaire, an AI Supercomputer, Toxic Emissions and a Memphis Community That Did Nothing Wrong • Tennessee Lookout.” *Tennessee Lookout*, 7 July 2025, https://tennesseelookout.com/2025/07/07/a-billionaire-an-ai-supercomputer-toxic-emissions-and-a-memphis-community-that-did-nothing-wrong/.

Brintrup, Alexandra, et al. “Trustworthy, Responsible, Ethical AI in Manufacturing and Supply Chains: Synthesis and Emerging Research Questions.” arXiv:2305.11581, arXiv, 19 May 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.2305.11581.

By. *Study: Industry Now Dominates AI Research | MIT Sloan*. 18 May 2023, https://mitsloan.mit.edu/ideas-made-to-matter/study-industry-now-dominates-ai-research.

Carnegie Endowment for International Peace. "China's AI Policy at the Crossroads: Balancing Development and Control in the DeepSeek Era." *Carnegie Endowment for International Peace*, 17 July 2025, https://carnegieendowment.org/research/2025/07/chinas-ai-policy-in-the-deepseek-era.

Cave, Stephen, and Seán S ÓhÉigeartaigh. "An AI Race for Strategic Advantage: Rhetoric and Risks." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* [New York, NY, USA], AIES '18, 2018, pp. 36–40, https://doi.org/10.1145/3278721.3278780.

Centre for Data Ethics and Innovation. "CDEI AI Barometer." *GOV.UK*, 23 June 2020, https://www.gov.uk/government/publications/cdei-ai-barometer/cdei-ai-barometer.

———. *Review into Bias in Algorithmic Decision-Making*. GOV.UK, 27 Nov. 2020.

ChatGPT, Google, Meta and Amazon: How Artificial Intelligence Is Really Powered | 7NEWS. https://7news.com.au/technology/chatgpt-google-meta-and-amazon-how-artificial-intelligence-is-really-powered--c-18992707. Accessed 10 Nov. 2025.

Chatterjee, Debashis, et al. "Optimizing Machine Learning for Water Safety: A Comparative Analysis with Dimensionality Reduction and Classifier Performance in Potability Prediction." *PLOS Water*, vol. 3, no. 8, Aug. 2024, p. e0000259. *PLoS Journals*, https://doi.org/10.1371/journal.pwat.0000259.

Chen, Brian J. "Great Power Antinomies." *Phenomenal World*, 16 Oct. 2025, https://www.phenomenalworld.org/analysis/great-power-antinomies/.

———. "Semiconductor Island." *Boston Review*, 16 Sept. 2024. *Boston Review*, https://www.bostonreview.net/articles/semiconductor-island/.

Chen, Xin, et al. "Electricity Demand and Grid Impacts of AI Data Centers: Challenges and Prospects." *arXiv.Org*, 8 Sept. 2025, https://arxiv.org/abs/2509.07218v3.

Chew Bigby, Bobbie. "Love of Place Over Lithium: Learning, Connecting, and Valuing Noongar Country." *Cultural Survival*, 13 Nov. 2023, https://www.culturalsurvival.org/news/love-place-over-lithium-learning-connecting-and-valuing-noongar-country.

Chung, Pyrou. *Indigenous Knowledge Systems and AI-Based Climate Action*. Digital Futures Lab, 2023, p. 17, https://www.climateai.asia/reports/DFL_AI_Issue_Brief1_IndigenousData_vF.pdf. AI + Climate Futures in Asia.

Cifas. *Fraudscape 2025 - Cifas*. 2025, https://www.fraudscape.co.uk/.

CISA. “Critical Infrastructure Sectors | CISA.” Web page / Government publication. *Critical Infrastructure Sectors*, https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors. Accessed 8 Nov. 2025.

Climate One. *Artificial Intelligence, Real Climate Impacts*. 19 Apr. 2024, https://www.climateone.org/audio/artificial-intelligence-real-climate-impacts.

Coburn, Jesse. “Government by AI? Trump Administration Plans to Write Regulations Using Artificial Intelligence.” *ProPublica*, 26 Jan. 2026, https://www.propublica.org/article/trump-artificial-intelligence-google-gemini-transportation-regulations.

Congress.gov. *Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress*. Legislation. 6 Apr. 2025, https://www.congress.gov/crs-product/R48555.

“Content Moderation Is a New Factory Floor of Exploitation — Labour….” *IHRB*, https://www.ihrb.org/latest/content-moderation-is-a-new-factory-floor-of-exploitation-labour-protections-must-catch-up. Accessed 10 Nov. 2025.

Contreras, Caesaro. “Do Autonomous Vehicles Deserve Your Trust? Experts Weigh In.” *Northeastern Global News*, 19 Dec. 2025, https://news.northeastern.edu/2025/12/19/waymo-automonous-vehicle-safety/.

Cooper, Sean Patrick. “‘The Precedent Is Flint’: How Oregon’s Data Center Boom Is Supercharging a Water Crisis.” *Rolling Stone*, 25 Nov. 2025, https://www.rollingstone.com/culture/culture-features/data-center-water-pollution-amazon-oregon-1235466613/.

“Counter Intelligence? Artificial Intelligence and Workers’ Health and Safety | LRD.” *Labour Research Department*, 2 Apr. 2025, https://www.lrd.org.uk/free-read/counter-intelligence-artificial-intelligence-and-workers-health-and-safety.

Cowls, Josh, et al. “The AI Gambit: Leveraging Artificial Intelligence to Combat Climate Change—Opportunities, Challenges, and Recommendations.” *AI & SOCIETY*, vol. 38, no. 1, Feb. 2023, pp. 283–307. *DOI.org (Crossref)*, https://doi.org/10.1007/s00146-021-01294-x.

Crawford, Kate. “Generative AI’s Environmental Costs Are Soaring — and Mostly Secret.” *Nature*, vol. 626, no. 8000, Feb. 2024, pp. 693–693. *DOI.org (Crossref)*, https://doi.org/10.1038/d41586-024-00478-x.

Crichton, Kyle; Ji, Jessica; Miller, Kyle; Bansemer, John; Arnold, Zachary; Batz, David; Choi, Minwoo; Decillis, Marisa; Eke, Patricia; Gerstein, Daniel M.; Leblang, Alex; McGee, Monty; Rattray, Greg; Richards, Luke; Scott, Alana. “Securing Critical Infrastructure in the Age of AI.” *Center for Security and Emerging Technology*, 2024, https://cset.georgetown.edu/publication/securing-critical-infrastructure-in-the-age-of-ai/.

Critical Entities Resilience Directive (CER) | Updates, Compliance, Training. https://www.critical-entities-resilience-directive.com/. Accessed 8 Feb. 2026.

Critical Infrastructure Sectors | CISA. https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors. Accessed 5 Nov. 2025.

Crownhart, Casey. “The Contentious Path to a Cleaner Future.” *MIT Technology Review*, 25 Jan. 2024, https://www.technologyreview.com/2024/01/25/1087086/nickel-mining-cleaner-future/.

Crum (Ben). *Brussels Effect or Experimentalism? The EU AI Act and Global Standard-Setting*. Info:eu-repo/semantics/article. 27 Aug. 2025, https://doi.org/10.14763/2025.3.2032.

Da, Longchao; Chen, Tiejin; Li, Zhuoheng; Bachiraju, Shreyas; Yao, Huaiyuan; Li, Li; Dong, Yushun; Hu, Xiyang; Tu, Zhengzhong; Wang, Dongjie; Zhao, Yue; Zhou, Xuanyu (Ben); Pendyala, Ram; Stabler, Benjamin; Yang, Yezhou; Zhou, Xuesong; Wei, Hua. “Generative AI in Transportation Planning: A Survey.” Preprint Archive (Open-Access). *arXiv.Org*, 2025, https://arxiv.org/html/2503.07158v4?ref=promptengineering.org&utm_source=chatgpt.com.

Darktrace. *The State of AI Cybersecurity*. Dec. 2025.

Das, Abhisek, et al. “AI Based Safety System for Employees of Manufacturing Industries in Developing Countries.” arXiv:1811.12185, arXiv, 28 Nov. 2018. *arXiv.org*, https://doi.org/10.48550/arXiv.1811.12185.

Dastin, Jeffrey, et al. “Paris AI Summit: France and EU Promise to Cut Red Tape on Tech.” *Reuters*, 10 Feb. 2025. Artificial Intelligence. *www.reuters.com*, https://www.reuters.com/technology/artificial-intelligence/paris-ai-summit-draws-world-leaders-ceos-eager-technology-wave-2025-02-10/.

“Data Science and AI Glossary.” *The Alan Turing Institute*, https://www.turing.ac.uk/news/data-science-and-ai-glossary. Accessed 1 Dec. 2025.

Davis, Peter Alexander Earls; Schmidt, Rebecca. “Standardised Bias? The Role — and Limits — of European Standards Bodies in the EU’s Artificial Intelligence Act | Internet Policy Review.” Online Open-Access Journal. *Internet Policy Review*, 2025, https://policyreview.info/articles/news/bias-european-standards-bodies.

Definitions. ICO, 9 July 2025, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/.

Department of Science, Technology and Innovation. “Future Risks of Frontier AI (Annex A).” *GOV.UK*, GOV.UK, 28 Apr. 2025, https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/future-risks-of-frontier-ai-annex-a.

Diemel, Caroline. “Understanding General Purpose AI.” *Eipa*, 11 Mar. 2025, https://www.eipa.eu/blog/understanding-general-purpose-ai/.

Domínguez Hernández, Andrés, et al. “Mapping the Individual, Social and Biospheric Impacts of Foundation Models.” *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* [New York, NY, USA], FAccT ’24, 2024, pp. 776–96. *ACM Digital Library*, https://doi.org/10.1145/3630106.3658939.

Dosunmu, Damilare, and Tessie Waithira. “Chinese Tech Companies Hire Kenyan Workers for AI Training - Rest of World.” *Rest of World*, https://restofworld.org/2025/kenya-china-ai-workers/. Accessed 18 Jan. 2026.

———. “The Hidden Kenyan Workers Training China’s AI Models.” *Rest of World*, 4 Dec. 2025, https://restofworld.org/2025/kenya-china-ai-workers/.

Dow, Jameson. “FSD False Advertising Case: Tesla Must Stop Lying or It Can’t Sell Cars, Judge Rules.” *Electrek*, 17 Dec. 2025, https://electrek.co/2025/12/17/ca-judge-rules-tesla-lied-about-fsd-must-fix-marketing-within-60-days/.

Dryer, Theodora. “Settler Computing: Water Algorithms and the Equitable Apportionment Doctrine on the Colorado River, 1950–1990.” *Osiris*, vol. 38, July 2023, pp. 265–85. *DOI.org (Crossref)*, https://doi.org/10.1086/725187.

Dunphy, Amy. “Is the Regulation of Connected and Automated Vehicles (CAVs) a Wicked Problem and Why Does It Matter?” *Computer Law & Security Review*, vol. 52, Apr. 2024, p. 105944. *ScienceDirect*, https://doi.org/10.1016/j.clsr.2024.105944.

E/ECE/TRANS/505/Rev.3/Add.151.

El-Helaly, Mohamed. “Artificial Intelligence and Occupational Health and Safety, Benefits and Drawbacks.” *La Medicina Del Lavoro*, vol. 115, no. 2, 2024, p. e2024014. *PubMed Central*, https://doi.org/10.23749/mdl.v115i2.15835.

Eubanks, Virginia. *Automating Inequality:* St Martin’s Press, 2017.

European Commission. *Article 26: Obligations of Deployers of High-Risk AI Systems | EU Artificial Intelligence Act*. 13 June 2024, https://artificialintelligenceact.eu/article/26/.

- *Article 60: Testing of High-Risk AI Systems in Real World Conditions Outside AI Regulatory Sandboxes | EU Artificial Intelligence Act*. 13 June 2024, https://artificialintelligenceact.eu/article/60/.
- *The General-Purpose AI Code of Practice | Shaping Europe’s Digital Future*. 2025, https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

European Commission, Directorate-General for Communications Networks, Content & Technology (DG CONNECT). “5G Observatory Report 2025 | Shaping Europe’s Digital Future.” *Shaping Europe’s Digital Future*, 2025, https://digital-strategy.ec.europa.eu/en/policies/5g-observatory-2025.

European Commission, Directorate-General for Migration and Home Affairs (DG HOME). “Critical Infrastructure Resilience at EU-Level - Migration and Home Affairs.” Web Page / Government Report. *European Commission — Home Affairs*, 2025, https://home-affairs.ec.europa.eu/policies/internal-security/counter-terrorism-and-radicalisation/protection/critical-infrastructure-resilience-eu-level_en.

European Parliament and Council of the European Union. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on Measures for a High Common Level of Cybersecurity across the Union, Amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and Repealing Directive (EU) 2016/1148 (NIS 2 Directive) (Text with EEA Relevance). *OJ L*, vol. 333, 14 Dec. 2022, http://data.europa.eu/eli/dir/2022/2555/oj/eng.

European Union. DIRECTIVE (EU) 2022/2555 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on Measures for a High Common Level of Cybersecurity across the Union, Amending Regulation (EU). 14 Dec. 2022.

European Union Agency for Cybersecurity (ENISA). "ENISA Threat Landscape 2024 | ENISA." *ENISA Threat Landscape 2024*, 6 Nov. 2025, https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024.

EVIR. *Self-Driving Cars and Electric Vehicles: U.S. Market Insights & Analysis, August - Electric Vehicle Intelligence Report*. 27 Aug. 2025, https://ev-intelligence.com/evs-self-driving-cars-aug2025/. Uncategorized.

Experian. *2026 Industry Data Breach Forecast*. 2026.

Feng, Jean, et al. "Clinical Artificial Intelligence Quality Improvement: Towards Continual Monitoring and Updating of AI Algorithms in Healthcare." *Npj Digital Medicine*, vol. 5, no. 1, May 2022, pp. 1–9. *www.nature.com*, https://doi.org/10.1038/s41746-022-00611-y.

FERN. *EU Use of Critical Raw Materials — How to Make Sure It's Fair for People and Forests*. 16 Mar. 2023, https://www.fern.org/publications-insight/eu-use-of-critical-raw-materials-how-to-make-sure-its-fair-for-people-and-forests/.

Fernández Llorca, David, et al. "Testing Autonomous Vehicles and AI: Perspectives and Challenges from Cybersecurity, Transparency, Robustness and Fairness." *European Transport Research Review*, vol. 17, no. 1, July 2025, p. 38. *BioMed Central*, https://doi.org/10.1186/s12544-025-00732-x.

Foy, Kylie. "New Tools Are Available to Help Reduce the Energy That AI Models Devour." *MIT News | Massachusetts Institute of Technology*, 5 Oct. 2023, https://news.mit.edu/2023/new-tools-available-reduce-energy-that-ai-models-devour-1005.

Gajjar, Devyani. *Artificial Intelligence (AI) Glossary*. Jan. 2024. *post.parliament.uk*, https://post.parliament.uk/artificial-intelligence-ai-glossary/.

Gandikota, Venkata. *2025 Accelerating a Frugal AI Ecosystem WhitePaper*. Nov. 2025.

Garcia-Lopez, Yvan J., et al. "Microfinance Institutions Failure Prediction in Emerging Countries, a Machine Learning Approach." *PLOS ONE*, vol. 20, no. 4, Apr. 2025, p. e0321989. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0321989.

Garofalo, Livia, Maia Woluchem, et al. "Digital Infrastructures, Material Consequences." *Data & Society*, 5 Feb. 2025, https://datasociety.net/points/digital-infrastructures-material-consequences/.

Garofalo, Livia, Joan Mukogosi, et al. "In Pennsylvania, a Nuclear Revival for an Uncertain AI Future." *Data & Society*, 20 Aug. 2025, https://datasociety.net/points/in-pennsylvania-a-nuclear-revival-for-an-uncertain-ai-future/.

Gerhold, Lars, and Edda Brandes. "Sociotechnical Imaginaries of a Secure Future." *European Journal of Futures Research*, vol. 9, no. 1, June 2021, p. 7. *BioMed Central*, https://doi.org/10.1186/s40309-021-00176-1.

Gestoso, Patricia. "Sustainable AI." *The Mint*, 5 July 2025, https://www.themintmagazine.com/sustainable-ai/.

Giambertoni, Marzia. *At the Paris AI Summit, Europe Charts Its Course*. 28 Feb. 2025. *www.rand.org*, https://www.rand.org/pubs/commentary/2025/02/at-the-paris-ai-summit-europe-charts-its-course.html.

Giannini, Alice, and Jonathan Kwik. "Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles." *Criminal Law Forum*, vol. 34, no. 1, Mar. 2023, pp. 43–85. *Springer Link*, https://doi.org/10.1007/s10609-023-09451-1.

Gihleb, Rania, et al. "Industrial Robots, Workers' Safety, and Health." *Labour Economics*, vol. 78, Oct. 2022, p. 102205. *ScienceDirect*, https://doi.org/10.1016/j.labeco.2022.102205.

Gilbey, Jess. "UK Construction Workers Embrace AI and Automation for Workplace Safety." *COSAC*, 29 May 2025, https://cosac.co.uk/news/uk-construction-workers-embrace-ai-and-automation-for-workplace-safety/.

Gillings. *Engineering Responsible AI: Foundations for Environmentally Sustainable AI*. Royal Academy of Engineering, 13 Dec. 2024.

Global Legal Insights. *AI, Machine Learning & Big Data Laws 2025 | China*. 15 May 2025, https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/china/.

Gooding, Matthew. *Virginia Narrowly Avoided Power Cuts When 60 Data Centers Dropped off the Grid at Once*. 20 Mar. 2025, https://www.datacenterdynamics.com/en/news/virginia-narrowly-avoided-power-cuts-when-60-data-centers-dropped-off-the-grid-at-once/.

Google. *Environmental Report 2024*. 2024.

Gornet (Mélanie) and Maxwell (Winston). *The European Approach to Regulating AI through Technical Standards*. Info:eu-repo/semantics/article. 16 July 2024, https://doi.org/10.14763/2024.3.1784.

Government Digital Service. "AI Insights: Agentic AI (HTML)." *GOV.UK*, 2025, https://www.gov.uk/government/publications/ai-insights/ai-insights-agentic-ai-html.

Government Office of Science. *Future Risks of Frontier AI: Which Capabilities and Risks Could Emerge at the Cutting Edge of AI in the Future*. Oct. 2023.

Goyal, Komal, et al. "Adoption of Artificial Intelligence-Based Credit Risk Assessment and Fraud Detection in the Banking Services: A Hybrid Approach (SEM-ANN)." *Future Business Journal*, vol. 11, no. 1, Mar. 2025, p. 44. *BioMed Central*, https://doi.org/10.1186/s43093-025-00464-3.

Gray, Mary L, and Siddarth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, 2019.

Gray, Mary L, and Siddharth Suri. *How to Stop Silicon Valley from Building a New Global Underclass*.

Greenblatt, Ryan, Denison, Carson, Wright, Benjamin, Roger, Fabian, MacDiarmid, Monte, Marks, Sam, Treutlein, Johannes, Belonax, Tim, Chen, Jack, Duvenad, David, Khan, Akbir, Michael, Julian, Minderman, Sören, Perez, Ethan, Petrini, Libra, Uesoto, Jonath, Kaplan, Jared, Shlegeris, Buck, Bowman, Samuel. R, Hubinger, Evan, "Alignment Faking in Large Language Models." *arXiv.Org*, 18 December 2024, https://arxiv.org/abs/2412.14093

Green Screen Coalition, et al. *Within Bounds: Limiting AI's Environmental Impact*. 5 Feb. 2025, https://greenscreen.network/en/blog/within-bounds-limiting-ai-environmental-impact/.

Greenstein, Stanley, and Mauro Zamboni. "Full Article: Navigating the Legislative Dilemma: Evaluating the EU AI Act's Approach to Regulating Emerging Technologies." *The Theory and Practice of Legislation*, vol. 13, 06 2025, https://www.tandfonline.com/doi/full/10.1080/20508840.2025.2513177.

Gregory, Andrew. "'Dangerous and Alarming': Google Removes Some of Its AI Summaries after Users' Health Put at Risk." *The Guardian*, 11 Jan. 2026. Technology. *The Guardian*, https://www.theguardian.com/technology/2026/jan/11/google-ai-overviews-health-guardian-investigation.

Gualtieri, Luca, et al. "Emerging Research Fields in Safety and Ergonomics in Industrial Collaborative Robotics: A Systematic Literature Review." *Robotics and Computer-Integrated Manufacturing*, vol. 67, Feb. 2021, p. 101998. *ScienceDirect*, https://doi.org/10.1016/j.rcim.2020.101998.

Guo, Eileen. "The Limits of Ethical AI." *Lighthouse Reports*, June 2025, https://www.lighthousereports.com/investigation/the-limits-of-ethical-ai/.

Gupta, Ashutosh. "Data Labeling for Autonomous Vehicles: The Road to Safe Automation." *Macgence AI*, 13 Jan. 2026, https://macgence.com/blog/data-labeling-for-autonomous-vehicles/.

Gutelius, Beth, and Nik Theodore. "KENYA'S DIGITALFIRST RESPONDERS: The Hidden Workforce Powering Global Tech." Center for Urban Economic Development University of Illinois Chicago, USA, Aug. 2025, https://cued.uic.edu/wp-content/uploads/sites/219/2025/08/cued_kdfr_final-1.pdf.

Gutfraind, Alexander, and Vicki Bier. "From Nuclear Safety to LLM Security: Applying Non-Probabilistic Risk Management Strategies to Build Safe and Secure LLM-Powered Systems." *arXiv.Org*, 20 May 2025, https://arxiv.org/abs/2505.17084v1.

Gwagwa, Arthur, et al. "Road Map for Research on Responsible Artificial Intelligence for Development (AI4D) in African Countries: The Case Study of Agriculture." *Patterns*, vol. 2, no. 12, Dec. 2021, p. 100381. *ScienceDirect*, https://doi.org/10.1016/j.patter.2021.100381.

Hameed, Mohammed Majeed, et al. "Forecasting Monthly Runoff in a Glacierized Catchment: A Comparison of Extreme Gradient Boosting (XGBoost) and Deep Learning Models." *PLOS ONE*, vol. 20, no. 5, May 2025, p. e0321008. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0321008.

Han, Shanshan. "Bridging Today and the Future of Humanity: AI Safety in 2024 and Beyond." *arXiv.Org*, 9 Oct. 2024, https://arxiv.org/abs/2410.18114v5.

Hao, Karen. *Empire of AI*. Penguin, 2025.

Heeks, Richard, et al. "China's Digital Expansion in the Global South: Systematic Literature Review and Future Research Agenda." *The Information Society*, vol. 40, no. 2, Mar. 2024, pp. 69–95. *Taylor and Francis+NEJM*, https://doi.org/10.1080/01972243.2024.2315875.

Hernandez Aros, Ludivia, et al. "Financial Fraud Detection through the Application of Machine Learning Techniques: A Literature Review." *Humanities and Social Sciences Communications*, vol. 11, no. 1, Sept. 2024, p. 1130. *www.nature.com*, https://doi.org/10.1057/s41599-024-03606-0.

Hernández, Jaime Galán. "ERT and AI: From Strategy to Action After the Paris Summit." *Telefónica*, 21 Mar. 2025, https://www.telefonica.com/en/communication-room/blog/ert-ai-strategy-action-paris-summit/.

———. "ERT and AI: From Strategy to Action After the Paris Summit." *Telefónica*, 21 Mar. 2025, https://www.telefonica.com/en/communication-room/blog/ert-ai-strategy-action-paris-summit/.

Hess, Julia Christina. *Chip Production's Ecological Footprint: Mapping Climate and Environmental Impact*. 20 June 2024, https://www.interface-eu.org/publications/chip-productions-ecological-footprint#acknowledgements.

High-Level Summary of the AI Act | EU Artificial Intelligence Act. https://artificialintelligenceact.eu/high-level-summary/. Accessed 8 Feb. 2026.

Hilliger, Laura, et al. *Harnessing AI for Environmental Justice*. 10 Feb. 2025, https://policy.friendsoftheearth.uk/reports/harnessing-ai-environmental-justice.

Hoffman, Bucknall, Benjamin. "Is Slop A.I. 's Answer to Spam? A Phrase Emerges for Bad Search. - The New York Times." *New York Times*, 11 June 2024, https://www.nytimes.com/2024/06/11/style/ai-search-slop.html.

Holzman, Jael. “How the Tech Industry Is Responding to Data Center Backlash - Heatmap News.” *Heatmap*, 16 Jan. 2026, https://heatmap.news/plus/the-fight/spotlight/data-center-backlash-response.

Hosseini, Mohammad, et al. “A Social-Environmental Impact Perspective of Generative Artificial Intelligence.” *Environmental Science and Ecotechnology*, vol. 23, Jan. 2025, p. 100520. *DOI.org (Crossref)*, https://doi.org/10.1016/j.ese.2024.100520.

“How AI Could Shape the Future of Health and Safety | NFP UK.” *NFP*, https://www.nfp.co.uk/media/insights/how-ai-could-shape-the-future-of-health-and-safety/. Accessed 24 Nov. 2025.

How Do Companies Protect Content Moderation Employees in Trust and Safety? 11 Sept. 2025, https://gearinc.com/protect-employees-involved-in-content-moderation/. Our Blogs.

Howey, William. “EU Acts to Secure Access to Critical Raw Materials.” *Economist Intelligence Unit*, 17 Apr. 2023, https://www.eiu.com/n/eu-acts-to-secure-access-to-critical-raw-materials/.

HR:Evolution — Automation Through the Ages. Directed by Paycom, 2022. *YouTube*, https://www.youtube.com/watch?v=g1OZDMcdm8Y.

“HSE Defends Use of AI for Protecting Workers at AGM.” *British Safety Council*, https://www.britsafe.org/safety-management/2025/hse-defends-use-of-ai-for-protecting-workers-at-agm. Accessed 19 Nov. 2025.

HSE’s Regulatory Approach to Artificial Intelligence (AI) — News - HSE. https://www.hse.gov.uk/news/hse-ai.htm. Accessed 24 Nov. 2025.

https://comparia.beta.gouv.fr/. https://comparia.beta.gouv.fr/. Accessed 9 Dec. 2025.

Ing, Lili Yan, and Gene M. Grossman. “Introduction.” *Robots and AI*, by Lili Yan Ing and Gene M. Grossman, 1st ed., Routledge, 2022, pp. 1–14. *DOI.org (Crossref)*, https://doi.org/10.4324/9781003275534-1.

Irani, Lilly. “The Hidden Faces of Automation.” *XRDS*, vol. 23, no. 2, Dec. 2016, pp. 34–37. *ACM Digital Library*, https://doi.org/10.1145/3014390.

Irving, Geoffrey. "Safety Cases at AISI | AISI Work." *AI Security Institute*, 23 Aug. 2023, https://www.aisi.gov.uk/blog/safety-cases-at-aisi.

Javaid, Mohd, et al. "Artificial Intelligence Applications for Industry 4.0: A Literature-Based Study." *Journal of Industrial Integration and Management*, vol. 07, no. 01, Mar. 2022, pp. 83–111. *DOI.org (Crossref)*, https://doi.org/10.1142/S2424862221300040.

Jaźwińska, Klaudia, and Aisvarya Chandrasekar. "AI Search Has a Citation Problem." *Columbia Journalism Review*, 6 Mar. 2025, https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php.

Kalai, Adam Tauman, et al. "Why Language Models Hallucinate." arXiv:2509.04664, arXiv, 4 Sept. 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2509.04664.

Kamran, Rashmi, et al. "Energy-Aware 6G Network Design: A Survey." arXiv:2509.11289, arXiv, 14 Sept. 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2509.11289.

Karankar, Nilima, and Anita Seth. "An IoT System for Access Control Using Blockchain and Message Queuing System." *EURASIP Journal on Information Security*, vol. 2025, no. 1, Oct. 2025, p. 31. *BioMed Central*, https://doi.org/10.1186/s13635-025-00208-4.

Kariuki, Njenga. *Economy | The 2025 AI Index Report | Stanford HAI*. 2025, https://hai.stanford.edu/ai-index/2025-ai-index-report/economy.

Keller, Anat, et al. "The European Union's Approach to Artificial Intelligence and the Challenge of Financial Systemic Risk." *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, edited by Henrique Sousa Antunes et al., Springer International Publishing, 2024, pp. 415–39. *Springer Link*, https://doi.org/10.1007/978-3-031-41264-6_22.

"Kenya Labor Court Rules That Facebook Can Be Sued." *AP News*, 6 Feb. 2023, https://apnews.com/article/technology-kenya-nairobi-business-mental-health-83b21e1c5a058f4221699dfc0cd16aa8.

Kerche, Francisco W., et al. "The Silicon Gaze: A Typology of Biases and Inequality in LLMs through the Lens of Place." *Platforms & Society*, vol. 3, Jan. 2026, p. 29768624251408919. *SAGE Journals*, https://doi.org/10.1177/29768624251408919.

Kerr, Dara. “AI Brings Soaring Emissions for Google and Microsoft, a Major Contributor to Climate Change.” *NPR*, 12 July 2024. Business. *NPR*, https://www.npr.org/2024/07/12/g-s1-9545/ai-brings-soaring-emissions-for-google-and-microsoft-a-major-contributor-to-climate-change.

Khanum, Noor ul Misbah, et al. “An Overview of the Prospects and Challenges of Using Artificial Intelligence for Energy Management Systems in Microgrids.” *arXiv.Org*, 6 May 2025, https://arxiv.org/abs/2505.05498v2.

Khurram, Minahil, et al. “Artificial Intelligence in Manufacturing Industry Worker Safety: A New Paradigm for Hazard Prevention and Mitigation.” *Processes*, vol. 13, no. 5, May 2025, p. 1312. *www.mdpi.com*, https://doi.org/10.3390/pr13051312.

Kierans, Aidan; Rittichier, Kaley; Sonsayar, Utku; Ghosh, Avijit. “Catastrophic Liability: Managing Systemic Risks in Frontier AI Development.” Preprint Archive (Open-access repository). *arXiv.Org*, 2025, https://arxiv.org/html/2505.00616v2?utm_source=chatgpt.com.

Kneese, Tamara. “Measuring AI’s Environmental Impacts Requires Empirical Research and Standards | TechPolicy.Press.” *Tech Policy Press*, 12 Feb. 2024, https://techpolicy.press/measuring-ais-environmental-impacts-requires-empirical-research-and-standards.

Kneese, Tamara, and Maia Woluchem. *Data Centers Aren’t the Future of American Prosperity*. Policy Brief. Data & Society, 22 July 2025, p. 8, https://datasociety.net/wp-content/uploads/2025/07/Myths-of-AI-Data-Centers-Arent-the-Future-of-American-Prosperity.pdf.

Knowledge, Whose. *State of the Internet Languages Report*. 2022, https://internetlanguages.org/en/.

Kou, Gang, and Yang Lu. “FinTech: A Literature Review of Emerging Financial Technologies and Applications.” *Financial Innovation*, vol. 11, no. 1, Jan. 2025, p. 1. *BioMed Central*, https://doi.org/10.1186/s40854-024-00668-6.

Lawrence, Alex. “Earning an ‘F’: AI Security Risks and The Telecoms Implications.” *TelcoForge*, 18 Aug. 2025, https://telcoforge.com/news/ai/new-reports-ai-security-risks-the-implications-for-telco/.

(Legislators), European Union. “Recital 55 - Classification of High-Risk AI Systems in Critical Infrastructure.” *AI Act*, 2022, https://ai-act-law.eu/recital/55/.

Lehuedé, Sebastián. "Territories of Data: Ontological Divergences in the Growth of Data Infrastructure." *Tapuya: Latin American Science, Technology and Society*, vol. 5, no. 1, Dec. 2022, p. 2035936. *Taylor and Francis+NEJM*, https://doi.org/10.1080/25729861.2022.2035936.

Letmathe, Peter, and Maren Paegert. "Automated Vehicles and Sustainability When Considering Rebound Effects." *PLOS ONE*, vol. 20, no. 8, Aug. 2025, p. e0329193. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0329193.

Li, Barbara. "Global AI Governance Law and Policy: China | IAPP." *IAPP.Org*, 12 Nov. 2025, https://iapp.org/resources/article/global-ai-governance-china.

Li, Biao, et al. "Editor's Introduction: Special Issue on the Anomie of AI in Finance; Financial Markets & Investments; Economic & Policy Analysis; Corporate Governance & Related Market Dynamics." *Financial Innovation*, vol. 11, no. 1, Sept. 2025, p. 121. *BioMed Central*, https://doi.org/10.1186/s40854-025-00792-x.

"Literature Review: Safe Adoption of Artificial Intelligence." *Google Docs*, https://docs.google.com/document/d/1SnJvO3ku9hrdmb75GnXCD-HFml6uP8AYEZfRgvfNK9I/edit?ouid=109961224193345836558&usp=docs_home&ths=true&usp=embed_facebook. Accessed 23 Jan. 2026.

Liu, Junhua. "A Survey of Financial AI: Architectures, Advances and Open Challenges." Preprint Archive (Open-access repository). *arXiv.Org*, 2024, https://arxiv.org/html/2411.12747v1?utm_source=chatgpt.com.

Liu, Qiren, et al. "Pain or Anxiety? The Health Consequences of Rising Robot Adoption in China." *Economics Letters*, vol. 236, Mar. 2024, p. 111582. *ScienceDirect*, https://doi.org/10.1016/j.econlet.2024.111582.

Liu, Shimiao. "LTR-Net: A Deep Learning-Based Approach for Financial Data Prediction and Risk Evaluation in Enterprises." *PLOS ONE*, vol. 20, no. 8, Aug. 2025, p. e0328013. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0328013.

Liu, Yang, et al. "Navigating Fintech and Banking Risks: Insights from a Systematic Literature Review." *Humanities and Social Sciences Communications*, vol. 12, no. 1, May 2025, p. 717. *www.nature.com*, https://doi.org/10.1057/s41599-025-05055-9.

Liu, Yizhi. *Worker-Centric Human-Robot Co-Adaptation in Construction*. 2023. The Pennsylvania State University, Ph.D. *ProQuest*, https://www.proquest.com/docview/3062065798/abstract/FFE35E00BE0142B3PQ/7.

Liu, YuanXiong, et al. “Understanding Perceived Ride Safety and Trust Formation in Robotaxi Services under Day and Night Conditions.” *Scientific Reports*, vol. 15, no. 1, Nov. 2025, p. 41798. *www.nature.com*, https://doi.org/10.1038/s41598-025-25722-w.

Luccioni, Sasha, et al. “Power Hungry Processing: Watts Driving the Cost of AI Deployment?” *The 2024 ACM Conference on Fairness Accountability and Transparency* [Rio de Janeiro Brazil], 2024, pp. 85–99. *DOI.org (Crossref)*, https://doi.org/10.1145/3630106.3658542.

Luke Kehoe. “Revealing the Cascading Impacts of the AWS Outage | Ookla®.” *Ookla - Providing Network Intelligence to Enable Modern Connectivity*, 22 Oct. 2025, https://www.ookla.com/articles/aws-outage-q4-2025.

Macapagal, Katrina. “The New Frontier: Managing and Insuring Generative and Agentic AI Risks.” *Edinburgh Futures Institute*, 20 Nov. 2025, https://efi.ed.ac.uk/the-new-frontier-managing-and-insuring-generative-and-agentic-ai-risks/.

Makulilo, A. B. “Privacy and Data Protection in Africa: A State of the Art.” *International Data Privacy Law*, vol. 2, no. 3, Aug. 2012, pp. 163–78. *DOI.org (Crossref)*, https://doi.org/10.1093/idpl/ips014.

Marin, Lucas G. Uberti-Bona, et al. “Are Companies Taking AI Risks Seriously? A Systematic Analysis of Companies’ AI Risk Disclosures in SEC 10-K Forms.” arXiv:2508.19313, arXiv, 28 Aug. 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2508.19313.

Markopoulou, Dimitra, and Vagelis Papakonstantinou. “The Regulatory Framework for the Protection of Critical Infrastructures against Cyberthreats: Identifying Shortcomings and Addressing Future Challenges: The Case of the Health Sector in Particular.” *Computer Law & Security Review*, vol. 41, July 2021, p. 105502. *ScienceDirect*, https://doi.org/10.1016/j.clsr.2020.105502.

Martinez-Velasco, Juan A., et al. “Survey on Methods for Detection, Classification and Location of Faults in Power Systems Using Artificial Intelligence.” arXiv:2507.10011, arXiv, 15 July 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2507.10011.

Masinde, Muthoni. “An Innovative Drought Early Warning System for Sub-Saharan Africa: Integrating Modern and Indigenous Approaches.” *African Journal of Science, Technology, Innovation and Development*, vol. 7, no. 1, Jan. 2015, pp. 8–25. *Taylor and Francis+NEJM*, https://doi.org/10.1080/20421338.2014.971558.

Massar, Moneim, et al. “Impacts of Autonomous Vehicles on Greenhouse Gas Emissions—Positive or Negative?” *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, May 2021, p. 5567. *PubMed Central*, https://doi.org/10.3390/ijerph18115567.

McQue, Katie, et al. “The Global Struggle over How to Regulate AI.” *Rest of World*, 21 Jan. 2025, https://restofworld.org/2025/global-ai-regulation-big-tech/.

McQuillann, Dan. *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*. Bristol University Press, 2022.

Milman, Oliver. “AI Is Guzzling Energy for Slop Content — Could It Be Reimagined to Help the Climate?” *The Guardian*, 17 Nov. 2025. Environment. *The Guardian*, https://www.theguardian.com/environment/2025/nov/17/ai-climate-crisis-cop30.

Mohamed, Shakir, et al. “Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence.” *Philosophy & Technology*, vol. 33, no. 4, Dec. 2020, pp. 659–84. *Springer Link*, https://doi.org/10.1007/s13347-020-00405-8.

Moorosi, Nyalleng. “Better Data Sets Won’t Solve the Problem — We Need AI for Africa to Be Developed in Africa.” *Nature*, vol. 636, no. 8042, Dec. 2024, pp. 276–276. *www.nature.com*, https://doi.org/10.1038/d41586-024-03988-w.

Moss, Emanuel, et al. *Assembling Accountability: Algorithmic Impact Assessment for the Public Interest*. Data & Society, 2021, https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/.

Mozilla.ai. “Mozilla.Ai - We’re Building a Future Where AI Works for You.” *Mozilla.Ai*, Jan. 2025, https://www.mozilla.ai.

Muldoon, James, et al. “The Poverty of Ethical AI: Impact Sourcing and AI Supply Chains.” *AI & SOCIETY*, vol. 40, Dec. 2023, pp. 529–43. *ResearchGate*, https://doi.org/10.1007/s00146-023-01824-9.

Muñoz-Carpena, Rafael, et al. “Convergence of Mechanistic Modeling and Artificial Intelligence in Hydrologic Science and Engineering.” *PLOS Water*, vol. 2, no. 8, Aug. 2023, p. e0000059. *PLoS Journals*, https://doi.org/10.1371/journal.pwat.0000059.

Murashov, Vladimir, et al. “Working Safely with Robot Workers: Recommendations for the New Workplace.” *Journal of Occupational and Environmental Hygiene*, vol. 13, no. 3, Mar. 2016, pp. D61–71. *PubMed Central*, https://doi.org/10.1080/15459624.2015.1116700.

Najem, Rihab, et al. “Advanced AI and Big Data Techniques in E-Finance: A Comprehensive Survey.” *Discover Artificial Intelligence*, vol. 5, no. 1, June 2025, p. 102. *Springer Link*, https://doi.org/10.1007/s44163-025-00365-y.

Navarro-Meneses, Francisco J., and Federico Pablo-Marti. “Reimagining Human Agency in AI-Driven Futures: A Co-Evolutionary Scenario Framework from Aviation.” *European Journal of Futures Research*, vol. 13, no. 1, Oct. 2025, p. 16. *BioMed Central*, https://doi.org/10.1186/s40309-025-00260-w.

Nelson, John P., et al. “Applications and Societal Implications of Artificial Intelligence in Manufacturing: A Systematic Review.” arXiv:2308.02025, arXiv, 25 July 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.2308.02025.

NHTSA. *Automated Driving Systems | NHTSA*. Text. https://www.nhtsa.gov/vehicle-manufacturers/automated-driving-systems. Accessed 22 Jan. 2026.

NIST. “AI Risk Management Framework.” *NIST*, Jan. 2023. *www.nist.gov*, https://www.nist.gov/itl/ai-risk-management-framework.

Nonnecke, Brandie, and Philip Dawson. *Human Rights Impact Assessments for AI: Analysis and Recommendations*. Access Now, 2022, p. 22, https://www.accessnow.org/wp-content/uploads/2022/11/Access-Now-Version-Human-Rights-Implications-of-Algorithmic-Impact-Assessments_-Priority-Recommendations-to-Guide-Effective-Development-and-Use.pdf.

Norwegian Consumer Council. *Ghost in the Machine: Addressing the Consumer Harms of Generative AI*. July 2023, p. 75, https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf.

O’Donnell, James, and Casey Crownhart. “We Did the Math on AI’s Energy Footprint. Here’s the Story You Haven’t Heard.” *MIT Technology Review*, 20 May 2025, https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/.

O’Hara, Kieron, and Wendy Hall. *Four Internets: Data, Geopolitics, and the Governance of Cyberspace*. Oxford University Press, 2021.

O’Kane, Sean. “Tesla’s Full Self-Driving Software under Investigation for Traffic Safety Violations.” *TechCrunch*, 9 Oct. 2025, https://techcrunch.com/2025/10/09/teslas-full-self-driving-software-under-investigation-for-traffic-safety-violations/.

Onat, Nuri C., et al. "Rebound Effects Undermine Carbon Footprint Reduction Potential of Autonomous Electric Vehicles." *Nature Communications*, vol. 14, no. 1, Oct. 2023, p. 6258. *www.nature.com*, https://doi.org/10.1038/s41467-023-41992-2.

O'Neill, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin, 2017.

O'Neill, Matt. *Data Centres in Ireland: The State of Play | IIEA*. The Institute of International and European Affairs, 2025, https://www.iiea.com/blog/data-centres-in-ireland-the-state-of-play.

Ookla. "Four Alarming Ways AI Weaponizes Vulnerabilities in Your Communication Infrastructure." *Ookla Research*, 2025, https://enterpriseconnect.com/four-alarming-ways-ai-weaponizes-vulnerabilities-in-your-communication-infrastructure/.

OSHA. *US Department of Labor Finds Amazon Exposed Workers to Unsafe Conditions, Ergonomic Hazards at Three More Warehouses in Colorado, Idaho, New York | Occupational Safety and Health Administration*. 2023, https://www.osha.gov/news/newsreleases/national/02012023.

PR Newswire. "Automation and AI Seen as Major Safety Opportunity but UK Workers Remain Unconvinced, Rapid Global Research Reveals." *Yahoo Finance*, 16 Oct. 2025, https://finance.yahoo.com/news/automation-ai-seen-major-safety-080000492.html.

Park, Ie Rei, et al. "Why Do People Resist AI-Based Autonomous Cars?: Analyzing the Impact of the Risk Perception Paradigm and Conditional Value on Public Acceptance of Autonomous Vehicles." *PLOS ONE*, vol. 20, no. 2, Feb. 2025, p. e0313143. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0313143.

Participation, Expert. “Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on Type-Approval Requirements for Motor Vehicles and Their Trailers, and Systems, Components and Separate Technical Units Intended for Such Vehicles, as Regards Their General Safety and the Protection of Vehicle Occupants and Vulnerable Road Users, Amending Regulation (EU) 2018/858 of the European Parliament and of the Council and Repealing Regulations (EC) No 78/2009, (EC) No 79/2009 and (EC) No 661/2009 of the European Parliament and of the Council and Commission Regulations (EC) No 631/2009, (EU) No 406/2010, (EU) No 672/2010, (EU) No 1003/2010, (EU) No 1005/2010, (EU) No 1008/2010, (EU) No 1009/2010, (EU) No 19/2011, (EU) No 109/2011, (EU) No 458/2011, (EU) No 65/2012, (EU) No 130/2012, (EU) No 347/2012, (EU) No 351/2012, (EU) No 1230/2012 and (EU) 2015/166 (Text with EEA Relevance) (Revoked).” Text. *Https://Webarchive.Nationalarchives.Gov.Uk/Eu-Exit/Https://Eur-Lex.Europa.Eu/Legal-Content/EN/TXT/?uri=CELEX:32019R2144*, King’s Printer of Acts of Parliament, https://www.legislation.gov.uk/eur/2019/2144. Accessed 22 Jan. 2026.

Pascal, Claire, et al. “Monitoring Indian Ungauged Small Reservoirs Volume from Remote Sensing: Feasibility, Bias and Perspectives.” *PLOS Water*, vol. 3, no. 12, Dec. 2024, p. e0000260. *PLoS Journals*, https://doi.org/10.1371/journal.pwat.0000260.

Pearse, Esme. “The Role of AI in Workplace Safety: A Critical Tool for Hazard Identification.” *AfterAthena*, 31 Mar. 2025, https://afterathena.co.uk/the-role-of-ai-in-workplace-safety-a-critical-tool-for-hazard-identification/.

Pelekis, Sotiris, et al. “Trustworthy Artificial Intelligence in the Energy Sector: Landscape Analysis and Evaluation Framework.” arXiv:2412.07782, arXiv, 12 Dec. 2024. *arXiv.org*, https://doi.org/10.48550/arXiv.2412.07782.

Peng, Bo, et al. “Artificial Intelligence in Human–Robot Collaboration in the Construction Industry: A Scoping Review.” *Buildings*, vol. 15, no. 17, Jan. 2025, p. 3060. *www.mdpi.com*, https://doi.org/10.3390/buildings15173060.

Perez, Jorge. *Tokenising Culture: Causes and Consequences of Cultural Misalignment in Large Language Models*. Ada Lovelace Institute, 19 June 2025, https://www.adalovelaceinstitute.org/blog/cultural-misalignment-llms/.

Porawagamage, Gayashan, et al. “A Review of Machine Learning Applications in Power System Protection and Emergency Control: Opportunities, Challenges, and Future Directions.” *Frontiers in Smart Grids*, vol. 3, Apr. 2024. *Frontiers*, https://doi.org/10.3389/frsgr.2024.1371153.

"Publications Office." https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2555. Accessed 8 Feb. 2026.

Raftree, Linda. "The US Won't Be Regulating AI Climate Impacts. Are There Sustainable AI Alternatives?" *MERL Tech*, 7 Nov. 2024, https://merltech.org/sustainable-ai-alternatives-to-big-tech-models/.

Raman, Raghu, et al. "Green and Sustainable AI Research: An Integrated Thematic and Topic Modeling Analysis." *Journal of Big Data*, vol. 11, no. 1, Apr. 2024, p. 55. *DOI.org (Crossref)*, https://doi.org/10.1186/s40537-024-00920-x.

Regattieri, Lori. *AI and Climate Change: The Global South Facing the New Geopolitics of Innovation*. Green Screen Coalition, 2025, p. 32, https://greenscreen.network/files/pdf/AI_and_Climate_Change-The_Global_South_Facing_The_New_Geopolitics_of_Innovation__EN_.pdf.

"Reimagining the Future of Data and AI Labor in the Global South." *Brookings*, https://www.brookings.edu/articles/reimagining-the-future-of-data-and-ai-labor-in-the-global-south/. Accessed 17 Nov. 2025.

Ren, Rui, et al. "Financial Risk Meter Based on Expectiles." SSRN Scholarly Paper no. 3809329, Social Science Research Network, 21 Mar. 2021. *papers.ssrn.com*, https://doi.org/10.2139/ssrn.3809329.

Revolutionizing Health and Safety: The Role of AI and Digitalization at Work | International Labour Organization. 22 Apr. 2025, https://www.ilo.org/publications/revolutionizing-health-and-safety-role-ai-and-digitalization-work.

Ricaurte Quijano, Paola. "Ethics for the Majority World: AI and the Question of Violence at Scale." *Media, Culture & Society*, vol. 44, no. 4, May 2022, pp. 726–45. *SAGE Journals*, https://doi.org/10.1177/01634437221099612.

Rismani, Shalaleh, et al. "From Silos to Systems: Process-Oriented Hazard Analysis for AI Systems." arXiv:2410.22526, arXiv, 29 Oct. 2024. *arXiv.org*, https://doi.org/10.48550/arXiv.2410.22526.

Roberts, Sarah T. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press, 2019.

Robison, Kylie. "Why Everyone Is Freaking out about DeepSeek." *The Verge*, 28 Jan. 2025, https://www.theverge.com/ai-artificial-intelligence/598846/deepseek-big-tech-ai-industry-nvidia-impac.

Robots and AI : A New Economic Era - EBSCO. https://research.ebsco.com/c/jfxa6m/ebook-viewer/pdf/ktfe5kfzdf/page/p_1. Accessed 24 Nov. 2025.

Rosenthal, Annie. "The Data Center Building Boom Is Running into Local Resistance." *Mother Jones*, 14 Sept. 2025, https://www.motherjones.com/politics/2025/09/data-center-construction-ai-water-local-resistance/. Environment.

Royal Commission into the Robodebt Scheme. 7 July 2023, https://robodebt.royalcommission.gov.au/.

Saha, Bikash, et al. "Generative AI in Financial Institution: A Global Survey of Opportunities, Threats, and Regulation." arXiv:2504.21574, arXiv, 1 May 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2504.21574.

Sarp, Salih, et al. "Digitalization of Railway Transportation through AI-Powered Services: Digital Twin Trains." *European Transport Research Review*, vol. 16, no. 1, Oct. 2024, p. 58. *BioMed Central*, https://doi.org/10.1186/s12544-024-00679-5.

Schütze, Paul. "The Impacts of AI Futurism: An Unfiltered Look at AI's True Effects on the Climate Crisis." *Ethics and Information Technology*, vol. 26, no. 2, June 2024, p. 23. *DOI.org (Crossref)*, https://doi.org/10.1007/s10676-024-09758-6.

Schwartz, Oscar. "Untold History of AI: How Amazon's Mechanical Turkers Got Squeezed Inside the Machine - IEEE Spectrum." *IEEE*, 2019, https://spectrum.ieee.org/untold-history-of-ai-mechanical-turk-revisited-tktkt.

Scott, Mark. "At Paris AI Summit, US, EU, Other Nations Lay Out Divergent Goals." *Tech Policy Press*, 11 Feb. 2025, https://techpolicy.press/at-paris-ai-summit-us-eu-other-nations-lay-out-divergent-goals.

Scroll. Click. Suffer: The Hidden Human Cost of Content Moderation and Data Labelling — Equidem. 2025, https://equidem.org/reports/scroll-click-suffer-the-hidden-human-cost-of-content-moderation-and-data-labelling/.

Sharma, Vivek, and Bhanu Priya. "Bridging the Gap: AI-Powered FinTech and Its Impact on Financial Inclusion and Financial Well-Being." *Discover Artificial Intelligence*, vol. 5, no. 1, Oct. 2025, p. 290. *Springer Link*, https://doi.org/10.1007/s44163-025-00465-9.

Shen, Judy Hanwen, and Alex Tamkin. "How AI Impacts Skill Formation." arXiv:2601.20245, arXiv, 1 Feb. 2026. *arXiv.org*, https://doi.org/10.48550/arXiv.2601.20245.

Shilongo, Kristophina. "A People-Centric Approach to AI in Africa Demands More Participation... From The People." *Tech Policy Press*, 25 Feb. 2025, https://techpolicy.press/a-people-centric-approach-to-ai-in-africa-demands-more-participation-from-the-people.

Shreya, and Harsh Pathak. "Explainable Artificial Intelligence Credit Risk Assessment Using Machine Learning." arXiv:2506.19383, arXiv, 25 June 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2506.19383.

Shu, Jun, et al. "Long-Term Water Demand Forecasting Using Artificial Intelligence Models in the Tuojiang River Basin, China." *PLOS ONE*, vol. 19, no. 5, May 2024, p. e0302558. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0302558.

Shumailov, Ilia, et al. "AI Models Collapse When Trained on Recursively Generated Data." *Nature*, vol. 631, no. 8022, July 2024, pp. 755–59. *www.nature.com*, https://doi.org/10.1038/s41586-024-07566-y.

Simone, Valentina De, et al. "Human-Robot Collaboration: An Analysis of Worker's Performance." *Procedia Computer Science*, vol. 200, 2022, pp. 1540–49. *DOI.org (Crossref)*, https://doi.org/10.1016/j.procs.2022.01.355.

Soffia, Magdalena, et al. *Impacts of Technology Exposure - a Report on Worker Wellbeing*. Mar. 2024. *DOI.org (Datacite)*, https://doi.org/10.5281/ZENODO.10470301.

Sofia. "Content Moderation and Data Labelling Map in Africa." *PersonalData.IO*, 10 Mar. 2025, https://personaldata.io/en/bpo-map/.

Sood, Aditya K, et al. "The Paradigm of Hallucinations in AI-Driven Cybersecurity Systems: Understanding Taxonomy, Classification Outcomes, and Mitigations." *Computers and Electrical Engineering*, vol. 124, May 2025, p. 110307. *ScienceDirect*, https://doi.org/10.1016/j.compeleceng.2025.110307.

SourceMaterial. "Big Tech's Data Centres Will Take Water from World's Driest Areas." *SourceMaterial*, 9 Apr. 2025, https://www.source-material.org/amazon-microsoft-google-trump-data-centres-water-use/.

Southern Environmental Law Center. *New Images Reveal Elon Musk's xAI Datacenter Has Nearly Doubled Its Number of Polluting, Unpermitted Gas Turbines - Southern Environmental Law Center*. 9 Apr. 2025, https://www.selc.org/press-release/new-images-reveal-elon-musks-xai-datacenter-has-nearly-doubled-its-number-of-polluting-unpermitted-gas-turbines/.

Spiegelhater, David. *The Art of Statistics: Learning from Data*. Pelican, 2019.

Staff, Entrepreneur. "Indo-Pacific Faces Growing AI Risks in Critical Infrastructure: Report." *Entrepreneur*, 27 Oct. 2025, https://www.entrepreneur.com/en-in/news-and-trends/indo-pacific-faces-growing-ai-risks-in-critical/498779.

Standard and Poor Global. "Beneath the Surface: Water Stress in Data Centers." *S&P Sustainable1*, 15 Sept. 2025, https://www.spglobal.com/sustainable1/en/insights/special-editorial/beneath-the-surface-water-stress-in-data-centers.

"Statement on AI Risk | CAIS." *Center for AI Safety*, 30 May 2023, https://aistatement.com.

Steingraber, Sandra, et al. "Data Centers and the Water Crisis." *The Science and Environmental Health Network*, 18 Aug. 2025, https://www.sehn.org/sehn/2025/8/14/data-centers-and-the-water-crisis.

Stern, Nicholas, et al. "Green and Intelligent: The Role of AI in the Climate Transition." *Npj Climate Action*, vol. 4, no. 1, June 2025, p. 56. *DOI.org (Crossref)*, https://doi.org/10.1038/s44168-025-00252-3.

Subramaniyan, Manochandar, et al. "Adaptive Resource Allocation and Routing for Integrated Sensing and Communications for Wireless Technologies." *EURASIP Journal on Wireless Communications and Networking*, vol. 2025, no. 1, May 2025, p. 33. *BioMed Central*, https://doi.org/10.1186/s13638-025-02461-0.

Sun, Kai, et al. "Improved MPC for Trajectory Planning of Self-Driving Cars." *PLOS ONE*, vol. 20, no. 6, June 2025, p. e0320359. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0320359.

Sun, Kejun; Liu, Shuo; Zhang, Yutian; Wang, Yijun. "Resilience-by-Design Concepts for 6G Communication Networks." Preprint Archive (Open-access Repository). *arXiv.Org*, 29 May 2024, https://arxiv.org/html/2405.17480v1?utm_source=chatgpt.com#abstract.

"Survey on Human–Robot Collaboration in Industrial Settings: Safety, Intuitive Interfaces and Applications | Request PDF." *ResearchGate*. *www.researchgate.net*, https://doi.org/10.1016/j.mechatronics.2018.02.009. Accessed 22 Nov. 2025.

Szadeczky, Tamas, and Zsolt Bederna. "Risk, Regulation, and Governance: Evaluating Artificial Intelligence across Diverse Application Scenarios." *Security Journal*, vol. 38, no. 1, June 2025, p. 35. *Springer Link*, https://doi.org/10.1057/s41284-025-00495-z.

Taft, Molly. *You're Thinking About AI and Water All Wrong | WIRED*. 12 Dec. 2025, https://www.wired.com/story/karen-hao-empire-of-ai-water-use-statistics/.

———. *You're Thinking About AI and Water All Wrong | WIRED*. 12 Dec. 2025, https://www.wired.com/story/karen-hao-empire-of-ai-water-use-statistics/.

Tamascelli, Nicola, et al. "Artificial Intelligence for Safety and Reliability: A Descriptive, Bibliometric and Interpretative Review on Machine Learning." *Journal of Loss Prevention in the Process Industries*, vol. 90, Aug. 2024, p. 105343. *ScienceDirect*, https://doi.org/10.1016/j.jlp.2024.105343.

Tan, Samson, et al. "The Risks of Machine Learning Systems." *arXiv.Org*, 21 Apr. 2022, https://arxiv.org/abs/2204.09852v1.

Tarkoma, Sasu, et al. "AI-Native Interconnect Framework for Integration of Large Language Model Technologies in 6G Systems." arXiv:2311.05842, arXiv, 10 Nov. 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.2311.05842.

Tessono, Blair Attard-Frost, Ana Brandusescu, David Gray Widder, Christelle. "AI Countergovernance: Lessons Learned from Canada and Paris." *Tech Policy Press*, 20 Feb. 2025, https://techpolicy.press/ai-countergovernance-lessons-learned-from-canada-and-paris.

Thapar, Ashish. "Asia Is Ahead of the Curve of Using AI to Fight Fraud. Here's What the Rest of the World Can Learn from It." *Fortune*, 31 Aug. 2025, https://fortune.com/asia/2025/08/31/asia-ai-bank-fraud-cybercrime-ntt-data/.

The AI Governance We Want, Call to Action: Liability, Interoperability, Sustainability & Labour. Internet Governance Forum, 2024, p. 111, https://intgovforum.org/en/filedepot_download/282/28491. Policy Network on Artificial Intelligence (PNAI).

The General-Purpose AI Code of Practice | Shaping Europe’s Digital Future. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai. Accessed 9 Dec. 2025.

THE GLOBAL HARMS OF POWERING ARTIFICIAL INTELLIGENCE. Directed by CPDPConferences, 2023. *YouTube*, https://www.youtube.com/watch?v=jrDrrAqJVbY.

The State Council Information Office, People’s Republic of China. *How Human-Robot Collaboration Is Powering China’s High-Quality Development | English.Scio.Gov.Cn*. Aug. 2025, http://english.scio.gov.cn/in-depth/2025-08/12/content_118022401.html.

The Turing Way. *The Environmental Impact of Digital Research*. 2023, https://book.the-turing-way.org/ethical-research/activism/activism-env-impact/.

Thomson, Gordon. “The Sovereign Critical Infrastructure Portfolio for Europe’s AI Future.” *Cisco News The EMEA Network*, 24 Sept. 2025, https://news-blogs.cisco.com/emea/2025/09/24/sovereign-critical-infrastructure-portfolio-for-europe-ai-future/.

Tian, Xu, et al. “Machine Learning in Internet Financial Risk Management: A Systematic Literature Review.” *PLOS ONE*, vol. 19, no. 4, Apr. 2024, p. e0300195. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0300195.

“Timeline History of Automation - How Automation Was Evolving.” *Progressive Automations*, 26 Apr. 2022, https://www.progressiveautomations.com/blogs/news/the-evolution-of-automation.

Tobey, Danny, et al. “China-Releases-AI-Safety-Governance-Framework.” *DLA Piper*, 12 Sept. 2024, https://www.dlapiper.com/en/insights/publications/2024/09/china-releases-ai-safety-governance-framework.

UK Examines AI in Workplace Safety. https://forms.zohopublic.eu/alaincharles/form/Pleaseenteryourdetails1/formperma/kzXzLVcI-2YpElFqnH6rQqM5k8x3iZGe0emGiQQh84k. Accessed 18 Nov. 2025.

UN News. *How AI Helps Combat Climate Change*. 3 Nov. 2023, https://news.un.org/en/story/2023/11/1143187.

UNCTAD. *Digital Economy Report 2024 Shaping an Environmentally Sustainable and Inclusive Digital Future*. United Nations, 2024, p. 288, https://unctad.org/system/files/official-document/der2024_en.pdf.

UNECE. *Agreement Concerning the Adoption of Harmonized Technical United Nations Regulations for Wheeled Vehicles, Equipment and Parts Which Can Be Fitted and/or Be Used on Wheeled Vehicles and the Conditions for Reciprocal Recognition of Approvals Granted on the Basis of These United Nations Regulations*. E/ECE/TRANS/505/Rev.3/Add.156, 14 Sept. 2017.

———. *Three Landmark UN Vehicle Regulations Enter into Force | UNECE*. 5 Feb. 2021, https://unece.org/sustainable-development/press/three-landmark-un-vehicle-regulations-enter-force.

UNEP. *Artificial Intelligence (AI) End-to-End*. Issues Note. 24 Sept. 2024, p. 6, https://wedocs.unep.org/bitstream/handle/20.500.11822/46288/AI-Environmental-Impact-Issues-Note.pdf?sequence=3&isAllowed=y.

UNEP and International Science Council. *Navigating New Horizons: A Global Foresight Report on Planetary Health and Human Wellbeing*. United Nations Environment Programme, 2024. *wedocs.unep.org*, https://wedocs.unep.org/xmlui/handle/20.500.11822/45890.

UrbanSDK. *How Different Countries Are Regulating Autonomous Vehicles*. https://www.urbansdk.com/resources/how-different-countries-are-regulating-autonomous-vehicles. Accessed 22 Jan. 2026.

U.S. Department of Homeland Security. "Groundbreaking Framework for the Safe and Secure Deployment of AI in Critical Infrastructure Unveiled by Department of Homeland Security | Homeland Security." *DHS News & Features (Archive)*, 2024, https://www.dhs.gov/archive/news/2024/11/14/groundbreaking-framework-safe-and-secure-deployment-ai-critical-infrastructure.

Utesch, Fabian, et al. "Towards Behaviour Based Testing to Understand the Black Box of Autonomous Cars." *European Transport Research Review*, vol. 12, no. 1, July 2020, p. 48. *BioMed Central*, https://doi.org/10.1186/s12544-020-00438-2.

van Wynsberghe, Aimee. "Sustainable AI: AI for Sustainability and the Sustainability of AI." *AI and Ethics*, vol. 1, no. 3, Aug. 2021, pp. 213–18. *Springer Link*, https://doi.org/10.1007/s43681-021-00043-6.

Varma, Girish, et al. "IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments." arXiv:1811.10200, arXiv, 26 Nov. 2018. *arXiv.org*, https://doi.org/10.48550/arXiv.1811.10200.

Varon, Joana, et al. *Fostering a Federated AI Commons Ecosystem*. Policy Brief. Brasil, 2024, p. 9, https://codingrights.org/docs/Federated_AI_Commons_ecosystem_T20Policybriefing.pdf. Task Force 05 Inclusive Digital Transformation.

Varoquaux, Gaël, et al. "Hype, Sustainability, and the Price of the Bigger-Is-Better Paradigm in AI." arXiv:2409.14160, arXiv, 1 Mar. 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2409.14160.

Vellinga, Nynke E. "Trustworthy AI on the Road." *Transport Transitions: Advancing Sustainable and Inclusive Mobility*, edited by Ciaran McNally et al., Springer Nature Switzerland, 2025, pp. 634–39. *Springer Link*, https://doi.org/10.1007/978-3-032-06763-0_91.

Vieira, Helena. "DeepSeek, ChatGPT and the Race towards Artificial General Intelligence - LSE Business Review." *LSE Business Review - Connecting Business Research with Policy, Practice and Public Debate*, 14 Mar. 2025, https://blogs.lse.ac.uk/businessreview/2025/03/14/deepseek-chatgpt-and-the-race-towards-artificial-general-intelligence/.

Vought, Russell T. *MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES*. Apr. 2025.

Walton, Jo Lindsay, et al. *The Cloud and the Climate: Navigating AI-Powered Futures*. Version 1.0, Sussex Digital Humanities Lab, 27 Sept. 2024. *DOI.org (Datacite)*, https://doi.org/10.5281/ZENODO.13850067.

"WEF_Navigating_the_AI_Frontier_2024.Pdf." https://reports.weforum.org/docs/WEF_Navigating_the_AI_Frontier_2024.pdf. Accessed 18 Jan. 2026.

What Are AI Hallucinations? | IBM. 1 Sept. 2023, https://www.ibm.com/think/topics/ai-hallucinations.

What It Will Take for India's New AI Governance Guidelines to Work | TechPolicy. Press. https://www.techpolicy.press/what-it-will-take-for-indias-new-ai-governance-guidelines-to-work/. Accessed 9 Dec. 2025.

Wheeler, Tom. "The Three Challenges of AI Regulation." *Brookings*, 15 June 2023, https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/.

White and Case LLP. *AI Watch: Global Regulatory Tracker - China | White & Case LLP*. 22 Sept. 2025, https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china.

———. *AI Watch: Global Regulatory Tracker - United States | White & Case LLP*. 24 Sept. 2025, https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states.

White House. America’s AI Action Plan. July 2025.

———. “Ensuring a National Policy Framework for Artificial Intelligence.” *The White House*, 11 Dec. 2025, https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/.

———. “Removing Barriers to American Leadership in Artificial Intelligence.” *The White House*, 23 Jan. 2025, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

“Who Is Responsible for Workplace Health and Safety?” *British Safety Council*, https://www.britsafe.org/training-and-learning/informational-resources/who-is-responsible-for-workplace-health-and-safety. Accessed 5 Nov. 2025.

Why Robots Are Harming Workers’ Mental Health - And What Companies Can Do About It - Bluesky Thinking. https://bluesky-thinking.com/why-robots-are-harming-workers-mental-health-and-what-companies-can-do-about-it/. Accessed 25 Nov. 2025.

Wikipedia contributors. “National Critical Information Infrastructure Protection Centre.” *Wikipedia*, 30 Sept. 2025. *Wikipedia*, https://en.wikipedia.org/w/index.php?title=National_Critical_Information_Infrastructure_Protection_Centre&oldid=1314287916.

Williams, Adrienne. *The Exploited Labor Behind Artificial Intelligence*. Oct. 2022. *www.noemamag.com*, https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence.

Willison, Simon. “Hallucinations in Code Are the Least Dangerous Form of LLM Mistakes.” Substack newsletter. *Simon Willison’s Newsletter*, 5 Mar. 2025, https://simonw.substack.com/p/hallucinations-in-code-are-the-least.

———. “Simon Willison: LLMs on Personal Devices.” *Simon Willison’s Weblog*, https://simonwillison.net/series/llms-on-personal-devices/. Accessed 8 Feb. 2026.

Winkle, Thomas, et al. "Area-Wide Real-World Test Scenarios of Poor Visibility for Safe Development of Automated Vehicles." *European Transport Research Review*, vol. 10, no. 2, June 2018, p. 32. *BioMed Central*, https://doi.org/10.1186/s12544-018-0304-x.

Wisakanto, Anna Katariina, et al. "Adapting Probabilistic Risk Assessment for AI." arXiv:2504.18536, arXiv, 2 July 2025. *arXiv.org*, https://doi.org/10.48550/arXiv.2504.18536.

Witt, Stephen. "Inside the Data Centers That Train A.I. and Drain the Electrical Grid." *The New Yorker*, 27 Oct. 2025. Brave New World Dept. *www.newyorker.com*, https://www.newyorker.com/magazine/2025/11/03/inside-the-data-centers-that-train-ai-and-drain-the-electrical-grid.

"Workplace Safety | UK Market Research Report — Rapid." *Https://Rapidglobal.Co.Uk/*, https://rapidglobal.co.uk/lp/workplace-safety-uk-market-research-report/. Accessed 24 Nov. 2025.

World Economic Forum. *Navigating the AI Frontier: A Primer on the Evolution and Impact of AI Agents*. Dec. 2024.

———. *The Global Risks Report 2026: 21st Edition*. 2026.

Xiao, Tianqi, et al. "Environmental Impact and Net-Zero Pathways for Sustainable Artificial Intelligence Servers in the USA." *Nature Sustainability*, Nov. 2025. *DOI.org (Crossref)*, https://doi.org/10.1038/s41893-025-01681-y.

Yang, Wenhao, et al. "Impact of Technological Advances on Workers' Health: Taking Robotics as an Example." *Sustainability*, vol. 17, no. 4, Jan. 2025, p. 1497. *www.mdpi.com*, https://doi.org/10.3390/su17041497.

York, University of. "Safe AI and Autonomous Systems Guidance." *University of York*, 2025, https://www.york.ac.uk/assuring-autonomy/guidance/.

Yu, Lining, et al. "An AI Approach to Measuring Financial Risk." arXiv:2009.13222, arXiv, 29 Sept. 2020. *arXiv.org*, https://doi.org/10.48550/arXiv.2009.13222.

Zaidan, Esmat, and Imad Antoine Ibrahim. "AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective." *Humanities and Social Sciences Communications*, vol. 11, no. 1, Sept. 2024, p. 1121. *www.nature.com*, https://doi.org/10.1057/s41599-024-03560-x.

Zander, Madelyn. *The Cloud Is Too Loud: Spotlighting the Voices of Community Activists from the Data Center Capital of the World*. 24 Aug. 2024, https://blog.castac.org/2024/08/the-cloud-is-too-loud-spotlighting-the-voices-of-community-activists-from-the-data-center-capital-of-the-world/.

Zewe, Adam. “Study: Transparency Is Often Lacking in Datasets Used to Train Large Language Models.” *MIT News | Massachusetts Institute of Technology*, 30 Aug. 2024, https://news.mit.edu/2024/study-large-language-models-datasets-lack-transparency-0830.

Zhang, Hui-Juan, et al. “Reliable Evaluation for the AI-Enabled Intrusion Detection System from Data Perspective.” *PLOS ONE*, vol. 20, no. 10, Oct. 2025, p. e0334157. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0334157.

Zou, Yueqing, and Yuanyuan Chen. “Hidden Cost of Automation: Do Industrial Robots Take a Toll on Our Mental Health?” *Journal of Asian Economics*, vol. 101, Dec. 2025, p. 102078. *ScienceDirect*, https://doi.org/10.1016/j.asieco.2025.102078.

February 2026

Lloyd’s Register Foundation
Report Series No: 2026.1